

Carla Alexandra Silva Ferreira

Preservação da Informação Digital: uma perspectiva orientada para as bibliotecas

2011



UNIVERSIDADE DE COIMBRA

DEPARTAMENTO DE FILOSOFIA, COMUNICAÇÃO E INFORMAÇÃO

FACULDADE DE LETRAS DA UNIVERSIDADE DE COIMBRA

Preservação da Informação Digital:

*uma perspectiva orientada para as
bibliotecas*

Carla Alexandra Silva Ferreira

Dissertação de Mestrado em *Informação, Comunicação e Novos Media*

apresentada à Faculdade de Letras da Universidade de Coimbra

sob a orientação da Professora Doutora Maria Manuel Borges

Coimbra

2011

Sumário

Agradecimentos	iii
Resumo	vii
<i>Abstract</i>	viii
Introdução	1
1. Preservação da informação na era digital	7
1.1. Valor da preservação digital	8
1.1.1. Preservação digital: uma missão transversal	9
1.2. Preservação digital: origens e definição do assunto	12
1.3. Preservação digital e as bibliotecas	17
1.3.1. Preservação tradicional vs Preservação digital	18
1.3.2. Digitalizar para preservar?.....	21
1.3.3. Processo de selecção em preservação digital.....	22
2. Documentos digitais: especificidade e desafios.....	25
2.1. Dependência e obsolescência tecnológica	25
2.2. Especificidade dos documentos digitais	29
2.3. Estrutura de um objecto digital.....	31
2.4. Propriedades significativas do objecto digital: o que preservar?.....	36
2.4.1. Autenticidade e integridade do objecto digital	36
3. Estratégias de preservação digital.....	41
3.1. Acções que preservam o objecto digital na sua dimensão física	45
3.1.1. Boas práticas de manuseamento e armazenamento	45
3.1.2. Refrescamento (<i>refreshing</i>)	46
3.1.3. Preservação da tecnologia (<i>tecnology preservation</i>)	48
3.2. Estratégias que preservam o objecto digital na sua dimensão lógica: <i>emulação</i>	49
3.3. Estratégias que preservam o objecto digital apenas na sua dimensão conceptual: <i>migração</i>	52
3.3.1. Migração <i>a-pedido</i>	56
3.3.2. Migração distribuída.....	58
3.3.3. Actualização de versões	59
3.3.4. Conversão para formatos concorrentes.....	60
3.4. Outras abordagens	61

3.4.1. Encapsulação (<i>Encapsulation</i>)	61
3.4.2. Arqueologia digital (<i>Digital archaeology</i>).....	62
3.4.4. Transferência para suportes analógicos.....	64
3.4.5. Não fazer nada.....	65
3.5. Escolha das estratégias de preservação digital	66
4. Repositórios digitais e a preservação digital	71
4.1. Modelo de referência OAIS	74
4.2. Metadados de preservação	78
5. Aspectos legais da preservação digital.....	85
5.1. Mecanismos de protecção como entraves à preservação digital	86
5.1.1. Protecção contratual	87
5.1.2. Protecção tecnológica.....	89
5.2. Depósito legal dos documentos digitais	90
5.3. Direitos de autor e a preservação digital: o caso português	93
5.3.1. O direito de autor e a preservação digital nos países da UE	99
Conclusão.....	103
Anexos	109
Anexo 1: <i>Boas práticas para o armazenamento e manuseamento de suportes magnéticos e suportes ópticos</i>	111
Anexo 2: <i>Factores de sustentabilidade, qualidade e funcionalidade dos formatos digitais</i>	113
Anexo 3: <i>Formatos adequados para preservação</i>	115
Anexo 4: <i>Diagrama esquemático para a selecção das estratégias de preservação digital</i>	119
Anexo 5: <i>Atributos e responsabilidades de um Repositório Digital de Confiança</i>	121
Anexo 6: <i>Modelo de referência OAIS – Pacotes de Informação</i>	123
Anexo 7: <i>Análise do direito de autor e a preservação digital nos países da UE</i>	127
Referências bibliográficas.....	131
Índice de figuras.....	141
Índice de tabelas.....	143

Agradecimentos

À minha orientadora, Doutora Maria Manuel Borges, da Universidade de Coimbra, pela orientação indispensável, pelo apoio incondicional e, sobretudo, pelo incentivo e amizade manifestados ao longo de todo este percurso.

Aos meus colegas de mestrado pelo companheirismo e pelo entusiasmo expressos no momento de partilha de angústias e incertezas. Aos meus amigos, pela amizade, claro está, e por terem contribuído para houvesse vida para além da dissertação.

À minha família, em particular, à mãe, exemplo de perseverança, e ao João Matela por, no culminar do cansaço, ter tido a paciência de rever o texto e sugerir melhorias.

Ao Tiago, por todas as razões que me dispenso de enumerar, mas sobretudo pelo carinho e pelo encorajamento. Sem ele – cuja presença foi constante, nesta fase tão especial da nossa vida – a execução deste trabalho não teria sido possível. A ele, expresso o meu profundo amor e gratidão.

Ao Francisco, fiel companheiro nesta jornada, que chegou antes de a dissertação estar terminada e que, por isso, com ela teve de partilhar colo.

Ao Malfamagrifinho

Resumo

A informação adquiriu, nas últimas décadas, uma enorme importância. Ela é um recurso social extremamente valioso, porque gera riqueza e contribui largamente para o progresso científico. Por força das novas tecnologias de informação e da *World Wide Web* - o novo meio de comunicação por excelência - o formato digital que a informação assumiu tornou mais fácil a sua produção, distribuição e disseminação, satisfazendo as crescentes necessidades das organizações e das pessoas. No entanto, por ser muito vulnerável, a informação digital coloca novos desafios em termos de preservação. A rápida degradação dos suportes que a veiculam e a obsolescência e dependência tecnológica tornam-na mais perecível. Por essa razão, considerar a preservação da informação digital é, por ventura, mais importante do que considerar a preservação da informação analógica, não só pelo valor económico que lhe está associado mas também porque a sua especificidade exige acções e estratégias de preservação contínuas e mais complexas. Apesar de a preservação da informação digital constituir uma missão transversal a várias entidades (implicando autores/produtores, produtores de *hardware* e *software*, editores/distribuidores, governos, utilizadores/consumidores, etc.), as bibliotecas continuam a ser os agentes insubstituíveis para garantir a perenidade do património cultural, literário e científico, agora em formato digital. Partindo desta premissa, o presente estudo apresenta os principais aspectos, conceitos, estratégias e desafios colocados pela preservação da informação digital, numa perspectiva orientada para as bibliotecas

PALAVRAS-CHAVE:

Preservação digital; bibliotecas; documentos digitais; estratégias de preservação digital; repositórios digitais; OAI, metadados de preservação; direitos de autor; depósito legal

Abstract

Information has gained, in the last decade, a huge importance. It is an extremely valuable social resource, for it generates wealth and it leads to scientific progress. Due to the new information technologies and the World Wide Web – the new privileged mean of communication – information has acquired a digital basis which has made its production, distribution and diffusion easier, fulfilling the growing needs of organizations and individuals. However, because it is too vulnerable, digital information presents new challenges when we think of its preservation. The fast degradation of the media and the technological obsolescence and dependence make this kind of information unlikely to endure. For that reason, thinking about the preservation of digital information is, perhaps, more important than considering the preservation of analogical information, not only because of its economical value but also because its uniqueness demands ongoing and more complex actions and preservation strategies. Although the preservation of digital information is a task for a multitude of entities (such as authors/producers, hardware and software engineers, publishers/distributors, governments, users/consumers, etc.), libraries are still the irreplaceable agents to ensure that the cultural, literary and scientific heritage remains imperishable, now in a digital basis. Assuming this thought as a starting point, this essay presents the main features, concepts, strategies and challenges placed by the preservation of digital information, through the libraries perspective.

KEYWORDS:

Digital preservation; libraries; digital documents; digital preservation strategies; digital repositories; OAIS, preservation metadata; copyright; legal deposit

Introdução

As sociedades contemporâneas são grandes produtoras e consumidoras de informação. O uso generalizado das tecnologias de informação e comunicação (TIC) motivou profundas alterações na forma como a informação passou a ser gerada, difundida, disseminada, armazenada e preservada. O impacto das novas tecnologias nas sociedades modernas é de tal ordem que está na origem de um tipo de sociedade em que a informação adquiriu um novo valor social, cultural e económico – a Sociedade da Informação.

A informação é entendida como um bem precioso, imprescindível para o desenvolvimento dos países e susceptível de gerar progresso e riqueza. As TIC, aliadas ao novo meio de comunicação por excelência, a *World Wide Web*, tornaram-se em ferramentas insubstituíveis para a criação, distribuição e consumo da informação e motivaram um aumento exponencial da sua produção e publicação. Pese embora a importância dos meios impressos (cujo volume paradoxalmente não diminui), a maior parte da informação passou a ser produzida digitalmente – com recurso a computadores, máquinas fotográficas, câmaras de vídeo digitais, etc. – e sem equivalente analógico.

Todas as esferas das sociedades modernas, sejam públicas ou privadas, são responsáveis pela criação de informação digital. A gestão e organização dessa informação têm afectado não só as tradicionais instituições de memória (arquivos, bibliotecas, centros de documentação, museus, etc.), mas também instituições como bancos, hospitais, empresas, entre outras.

São evidentes as vantagens e os atractivos dos documentos digitais relativamente aos documentos analógicos. Os primeiros podem agregar conteúdos multimédia e ocupam pouco volume (do ponto de vista da capacidade de armazenamento de informação, quando comparados com os equivalentes analógicos); a volatilização do suporte deslocaliza-os no tempo e no espaço, permitindo o acesso simultâneo de várias pessoas ao mesmo documento. Além disso, os documentos digitais são facilmente transmissíveis, replicáveis e editáveis, inerentemente pesquisáveis e, em alguns contextos, permitem estabelecer ligações entre vários conteúdos através de hiperligações (*links*).

No entanto, em virtude das suas especificidades, a preservação dos documentos digitais coloca novos desafios: o acesso aos documentos depende do ambiente tecnológico que permite reproduzi-los; a sua virtualidade e volatilidade obrigam a repensar aspectos como a autenticidade e integridade, os direitos de autor, a autoria, o controlo de versões, etc.

O novo contexto tecnológico obriga as instituições de memória (em particular, as bibliotecas) a repensarem a forma como preservam as suas colecções para poderem continuar a cumprir a sua missão de curadoras do património literário, científico e cultural agora em formato digital. Por essa razão, não é de estranhar que, sobretudo a partir da década de 80 do século XX, tenham sido as instituições de memória as primeiras a alertar para a necessidade de definir medidas e estratégias que permitissem preservar a informação digital a longo prazo, talvez por adivinharem que a inércia traria consigo cenários devastadores:

As we moves [sic] into the electronic era of digital objects, it is important to know that there are new barbarians at the gate and that we are moving into an era where much of what we know today, much of what is coded and written electronically, will be lost forever. We are, to my mind, living in the midst of digital Dark Ages; consequently, much as monks oftimes past, it falls to librarians and archivists to hold to the tradition which reveres history and the published heritage of our times. (Kuny, 1997, p. 1)

O repto às bibliotecas estava lançado e a necessidade de agir tornou-se evidente, sobretudo, a partir dos anos 90 do século passado. Os desafios suscitados pela preservação digital são distintos dos que a preservação tradicional coloca; além disso, as instituições e os agentes implicados nessa missão excedem as tradicionais instituições de memória. Tornou-se necessário, por isso, definir conceitos, identificar problemas e ameaças e delinear estratégias de actuação no âmbito da preservação digital. Foi no mundo anglo-saxónico, pioneiro nesta área de estudo, que surgiram os primeiros estudos e publicações sobre preservação digital. Paralelamente às bibliotecas dos Estados Unidos da América e da Grã-Bretanha, as bibliotecas da Europa e da Oceânia, sobretudo as universitárias, começam a implementar os primeiros projectos de preservação digital.

Dos inúmeros projectos¹ de preservação digital existentes resultaram vários estudos e publicações. Em Portugal, há conhecimento de alguns projectos² de preservação digital, mas a literatura sobre o tema permanece escassa. A recente publicação da tese de doutoramento de Miguel Ferreira – *Preservação de longa duração de informação digital no contexto de um arquivo histórico* (2009) – bem como a literatura publicada no âmbito desses projectos

¹ Pinto (2007) procedeu ao levantamento, mapeamento e análise de 665 programas e projectos de preservação digital em 18 países, entre 1995 e 2007.

² Destaque-se o projecto RODA (*Repositório de Objectos Digitais Autênticos*) lançado pela Direcção-Geral de Arquivos (DGARQ) /Torre do Tombo e que contou com a colaboração da Universidade do Minho. Este projecto foi concebido com a missão específica de recolher, preservar e facultar o acesso a património arquivístico digital, acompanhando os novos desenvolvimentos científicos e técnicos no domínio da preservação digital.

trouxeram um importante e valioso contributo para o estudo da preservação digital em Portugal. No entanto, estes estudos são, em grande parte, elaborados sob a perspectiva dos arquivos.

Ainda que o propósito seja o mesmo – preservar a informação digital – a missão de um arquivo é distinta da missão de uma biblioteca, pois é também diferente a natureza da informação que tutelam. Por essa razão, há aspectos da preservação digital que têm de ser perspectivados de forma distinta. Mesmo que as acções e as estratégias de preservação possam ser as mesmas, aspectos como as propriedades significativas, a autenticidade, a integridade e o valor cultural da informação num projecto de preservação digital levado a cabo por um arquivo podem ser abordados de forma diferente num projecto de preservação digital levado a cabo por uma biblioteca. Outro aspecto que pode ser distinto diz respeito às questões legais levantadas pela preservação da informação digital. O depósito legal, por exemplo, que num arquivo não tem qualquer razão de ser, é de suma importância para a preservação da informação digital numa biblioteca. Sem um depósito legal alargado aos documentos digitais, é muito difícil a preservação da produção cultural, literária e científica portuguesa em formato digital, pois as bibliotecas só podem preservar aquilo que possuem.

A parca existência de literatura em língua portuguesa sobre a preservação da informação digital em bibliotecas e a ausência de estudos/projectos sobre preservação digital nas bibliotecas portuguesas constitui, simultaneamente um estímulo e um desafio para a elaboração deste estudo. Por um lado, o incentivo de abordar um tema pouco investigado em Portugal (pelo menos na perspectiva das bibliotecas), por outro lado, o desafio que constitui elaborar um trabalho em que é manifesta a carência bibliográfica sobre o assunto em estudo. Perante a escassa publicação de literatura sobre preservação da informação digital numa perspectiva orientada para as bibliotecas, fica-se sem saber se as bibliotecas portuguesas³ possuem ou não um plano de preservação para as suas colecções digitais⁴.

O presente estudo tem como principal objectivo reflectir sobre a preservação da informação digital numa perspectiva orientada para as bibliotecas e pretende colmatar aquilo que se verificou ser um hiato na literatura portuguesa sobre esse assunto. Por essa razão, o

³ A Biblioteca Digital Nacional (que resulta de um projecto da Biblioteca Nacional de Portugal, lançado em 2002) oferece uma série de serviços técnicos de armazenamento, organização e apresentação dos objectos digitais e integra ainda serviços que suportam a sua utilização pública, propiciando diversas formas de acesso livre e gratuito aos documentos digitalizados. No entanto, não é claro o comprometimento com a preservação desses documentos a longo prazo, nem é pública a forma como eles serão reservados no futuro.

⁴ Borbinha (2004) dá a conhecer algumas iniciativas da BND: *DiTeD – Teses e Dissertações Digitais e Repositório de Preservação* e apresenta alguns dos problemas do depósito e preservação de informação digital na Biblioteca Nacional e a forma como a BND tem lidado com eles.

estudo é basicamente teórico e expositivo, em resultado de uma leitura e de uma análise crítica da principal literatura nacional⁵ e internacional⁶ sobre a preservação da informação digital publicada sobretudo na primeira década do século XXI. Verificou-se que a maior parte dos autores, bem como as revistas científicas onde são publicados os artigos, são de origem anglo-americana. Este facto motivou ainda mais a elaboração de um trabalho que pretende ser um documento orientador para as bibliotecas portuguesas. Porque foi concebido sob o ponto de vista da Ciência da Informação, mais do que uma abordagem técnica e tecnicista da preservação da informação digital, o estudo pretende especificamente:

- a) Contextualizar e definir a preservação da informação digital e o seu valor na sociedade actual;
- b) Reflectir sobre a missão de preservação das bibliotecas na era digital;
- c) Identificar as principais ameaças à informação digital e os principais desafios que esta coloca à preservação;
- d) Definir e sistematizar os principais conceitos relacionados com a preservação da informação digital;
- e) Enunciar e descrever as principais acções e estratégias de preservação da informação digital;
- f) Caracterizar as novas infra-estruturas técnicas e sociais, tais como os repositórios digitais, que visam a perenidade da informação digital;
- g) Abordar os principais aspectos legais relacionados com a preservação da informação digital e analisar a forma como a legislação portuguesa (im)possibilita a aplicação das principais estratégias de preservação.

O estudo apresenta-se dividido em cinco capítulos. O primeiro é um capítulo introdutório, onde se refere a importância da informação (sobretudo daquela que se encontra em formato digital) e o valor e utilidade da preservação digital na sociedade contemporânea. Define-se e contextualiza-se o tema em estudo e identificam-se os agentes implicados na preservação digital e as respectivas responsabilidades. Por fim, aborda-se a missão de preservação das bibliotecas à luz do novo contexto: identificam-se as principais diferenças entre

⁵ A pesquisa de literatura sobre preservação digital em Portugal foi feita na PORBASE (Catálogo Colectivo das Bibliotecas Portuguesas), no RAACP (Repositório Científico de Acesso Aberto em Portugal) e nas edições da APBAD (Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas).

⁶ A pesquisa de literatura sobre o tema foi feita, em grande medida, nas bases de dados *ISI Web of Knowledge* e *Biblioteca do Conhecimento on-line (B-on)*, sendo a selecção feita de acordo com a relevância e ano de publicação. Segundo a ISI, 2007 e anos envolventes (2005, 2006 e 2008) são o período em que mais se publicou sobre este assunto.

a preservação tradicional e a preservação digital, avalia-se o valor dos substitutos digitais e analisa-se sucintamente o processo de selecção de documentos digitais para preservação.

No segundo capítulo abordam-se a especificidade e os desafios colocados pela informação digital, decorrentes nomeadamente da sua volatilidade, da sua dependência tecnológica e da rápida degradação dos suportes que a veiculam. Este capítulo é essencial para a compreensão do capítulo seguinte, pois nele é analisada a estrutura dos objectos digitais e são definidos conceitos que permitem a interpretação do capítulo subsequente. São explicadas as dimensões física, lógica e conceptual dos objectos digitais⁷ e conceitos como propriedades significativas, autenticidade e integridade são interpretados à luz da preservação digital.

O terceiro capítulo versa sobre as principais acções e estratégias de preservação organizadas consoante o enfoque da preservação seja dado à dimensão física, lógica ou conceptual do objecto digital. Deste modo, são distinguidas: (i) as acções que preservam o objecto digital na sua dimensão física (boas práticas de armazenamento e manuseamento; refrescamento; preservação da tecnologia); (ii) as estratégias que preservam o objecto digital na sua dimensão lógica (emulação); (iii) as estratégias que preservam o objecto digital apenas na sua dimensão conceptual (migração); (iv) e outras abordagens (onde se incluiu acções que ou são transversais às abordagens anteriores, ou são menos viáveis ou pouco experimentadas). Este capítulo termina com algumas indicações de uso de cada acção e/ou estratégia de preservação, sendo também enunciadas algumas linhas orientadoras para selecção da melhor estratégia de preservação para dos diferentes tipos de recursos digitais.

O quarto capítulo deste trabalho aborda o paradigma actual da preservação digital, à luz do qual os repositórios surgem como as infra-estruturas técnicas e sociais por excelência que asseguram a perenidade dos documentos digitais. É definido repositório digital e distinguem-se as principais responsabilidades de um repositório digital de confiança. Naturalmente, o capítulo inclui a descrição (ainda que sucinta) do modelo de referência OAIS como estrutura-base para a criação de sistemas de preservação e gestão de informação digital. Outro aspecto essencial abordado neste capítulo são os metadados de preservação: eles surgem, no contexto da preservação digital, como a informação descritiva e técnica sobre os objectos digitais que não só documenta e suporta as actividades de preservação, mas também descreve e localiza os materiais digitais preservados, facilitando a sua pesquisa e o seu acesso no futuro.

⁷ Neste trabalho utilizam-se as expressões “documento digital”, “recurso digital”, “material digital” e “objecto digital” de forma indistinta e unívoca para designar objectos de informação codificada digitalmente (que pode assumir a forma de texto, base de dados, imagem, vídeo, som, material gráfico, programa informático, página *Web* ou aplicação de *software*) foco de preservação.

Finalmente, no quinto capítulo são evocados os aspectos legais relacionados com a preservação da informação digital: em concreto, os direitos de autor e o depósito legal. A análise da legislação portuguesa sobre estes dois aspectos legais é antecedida de uma breve exposição sobre os mecanismos de protecção utilizados pelos detentores do *copyright*. Numa tentativa de controlar o acesso e a distribuição da produção intelectual, muitos produtores e/ou detentores dos direitos de autor reforçaram os seus direitos através de mecanismos de protecção contratual e tecnológica, os quais, em muitos casos, entravam a aplicação das principais estratégias de preservação. A existência destes mecanismos de protecção – que podem também inibir o acesso à informação –, suscita a análise, neste último capítulo, da legislação portuguesa e europeia a fim de averiguar de que modo o quadro legal salvaguarda os privilégios e excepções de uso concedidos às bibliotecas – condição para que estas possam, em ambiente digital, preservar e tornar acessível a informação.

Todos os aspectos aqui enunciados serão sempre abordados numa perspectiva orientada para as bibliotecas e sob o ponto de vista da Ciência da Informação. Não fazem parte do âmbito deste estudo aspectos como planeamento e a gestão da preservação digital, embora eles sejam imprescindíveis a qualquer projecto de preservação digital. As directrizes da UNESCO sobre preservação digital oferecem informações detalhadas sobre a implementação e gestão de projectos de preservação digital e podem servir como documento orientador a quem pretenda desenvolver um projecto desse tipo.

A premissa básica deste estudo é que as bibliotecas, na era digital, devem continuar a cumprir a sua missão de preservação do património cultural, científico e literário, que cada vez mais assume o formato digital. Por esta razão, pretende-se que este trabalho sirva de documento introdutório e orientador, oferecendo aos bibliotecários uma leitura da preservação digital à luz da Ciência da Informação.

1. Preservação da informação na era digital

Cada vez mais o património cultural, artístico e científico se reveste de formas digitais. Estudos recentes revelam que nunca se produziu tanta informação como nas últimas décadas. Gladney (2007, p. 25 e 43) refere alguns números da *Long Now Foundation*⁸, que estima que, desde 1945, a Humanidade criou e armazenou cem vezes mais informação do que em toda a sua história até essa data. Prevê-se que nos próximos cinco anos a Humanidade produza mais informação – desde *web sites* a fotos e vídeos digitais – do que aquela que foi produzida nos últimos mil anos! O universo da informação digital está a crescer incontrolavelmente e contabilizá-lo é uma tarefa titânica. Um estudo sobre o universo digital, levado a cabo pela *International Data Corporation*⁹ (IDC), sob a direcção de John F. Grantz (2008), apoiado pela *EMC Corporation*¹⁰, mostra que o universo digital está cada vez maior e que o seu crescimento é cada vez mais rápido¹¹:



Figura 1: Informação digital criada, capturada e replicada no mundo inteiro. Fonte: Grantz (2008, p. 5)

⁸ Disponível na WWW: <http://www.longnow.org/>

⁹ Disponível na WWW: <http://www.idc.pt/index.html>

¹⁰ Disponível na WWW: <http://portugal.emc.com/>

¹¹ Em 2007, o universo digital, ou seja, a informação criada, capturada e replicada em formato digital, era da ordem dos 281 Exabytes (281 mil milhões de Gb). Em 2009, esse universo era de 773 Exabytes e prevê-se que, em 2011, atinja 1800 Exabytes. Se dividirmos os 281 mil milhões de Gb de informação digital pelos 6,6 mil milhões de habitantes do mundo em 2007, o rácio é de 43 Gb de informação por pessoa. A digitalização das 372 páginas da primeira edição de *Os Lusíadas* feita pela Biblioteca Nacional equivale a 27 Mb. Se cada um dos habitantes do mundo fosse responsável pela produção de 43Gb de informação, cada pessoa teria escrito 1630 livros como *Os Lusíadas* de Luís de Camões.

O ritmo a que a informação digital cresce obriga a não descurar o facto de a instabilidade e a vulnerabilidade dos documentos em formato digital colocarem sérios problemas de preservação. Se a cadência com que a produção da informação digital é produzida, não for acompanhada das acções e estratégias necessárias à sua preservação, a perda dos dados armazenados digitalmente¹² pode revelar-se catastrófica para a Humanidade.

1.1. Valor da preservação digital

A problemática da preservação digital ganha maior relevância por vivermos em plena Sociedade da Informação, cuja economia é pautada pelo acesso à informação. A revolução informacional que marca esta nova ordem social consiste numa mudança da forma como o conhecimento é comunicado: houve melhorias na rapidez e precisão do registo da informação, a sua transmissão passou a ser mais rápida e os sistemas de recuperação mais eficazes.

A informação digital é um recurso vital para todos os sectores da Sociedade da Informação, mas é também um recurso intrinsecamente frágil e efêmero, porque os seus suportes são instáveis e porque depende de um ambiente tecnológico em constante evolução.

A preservação da informação digital assume um elevado valor social: os benefícios que traz são incomensuráveis e transcendem os interesses individuais, particulares ou proprietários. O valor da informação reside na possibilidade de esta poder gerar conhecimento: este, por sua vez, pode ser aplicado na investigação e na educação (informação científica), nos negócios e nas indústrias (informação técnica), no quotidiano e na formação de cidadãos conscientes, críticos e activos (informação histórico-cultural)¹³. Deste modo, a informação digital deve ser entendida como um bem público, um benefício social e deve ser tratada como tal.

¹² Existem vários exemplos de perda de dados armazenados digitalmente. O mais citado é o do *U.S. Census Bureau* de 1960, cujos dados foram armazenados em cassetes UNIVAC. Quando, em 1976, se tentou aceder a estes dados, esse suporte tinha-se tornado completamente obsoleto, existindo no mundo apenas dois computadores capazes de o ler. Foram precisos esforços consideráveis para copiar 1,5 milhões de registos para novos suportes; todavia 10,000 desses registos não puderam ser copiados.

¹³ Smith (2007, p. 19) lembra que o controlo da informação é uma das estratégias dos regimes totalitários para controlar os seus cidadãos: “A people who do not own and control their own cultural heritage are a people who can be held captive by false histories, fabrications, and lies. The genius of totalitarian societies is that the need for brute force to make subjects out of citizens is really quite modest. If the government controls what people know about their past and their present, they limit the scope of their imaginations and can control their expectations for the future”. Precisamos do conhecimento histórico (conhecimento sobre o passado) e de memória (conhecimento do passado), para nos tornarmos plenos, autênticos e confiantes para que haja coesão social.

Ainda é difícil determinar o custo da preservação da informação digital a longo prazo. No entanto, sabendo-se que a possibilidade de perda irrecuperável do património digital¹⁴ causará custos irreparáveis, devem os governos investir no sentido de assegurar a sua perenidade, esclarecer as pessoas da importância social da preservação digital, tomar medidas e comprometer-se com políticas de preservação, ponderando os custos e a forma como estes deverão ser suportados, fomentando a colaboração entre os vários intervenientes no ciclo de vida de um documento digital, entre outras medidas.

1.1.1. Preservação digital: uma missão transversal

Preservar o património cultural, literário e científico digital é um dever que as gerações actuais têm para com as gerações futuras. Uma responsabilidade dessa magnitude deve ser universal e partilhada, implicando não apenas as instituições tradicionais mas também os produtores da informação digital, os editores e os governos.

Na *Carta para a Preservação do Património Digital da Unesco* (2003), os Estados-Membros surgem como os principais impulsores da preservação do património digital, cabendo-lhes (i) trabalhar em estreita colaboração com as organizações e instituições pertinentes de forma a empreender actividades de divulgação e promoção da preservação do património digital; (ii) formular políticas; (iii) sensibilizar o público em geral (desde os criadores e gestores de informação digital até aos seus consumidores) sobre os problemas práticos que este tipo de documentação coloca em termos de preservação; (iv) proporcionar um contexto jurídico que maximize a acessibilidade e autenticidade do património digital (alargando o depósito legal aos documentos digitais, promovendo um justo equilíbrio entre o direito de autor e o interesse público, prevenindo e protegendo o património digital de manipulação ou modificação, ...).

Uma das responsabilidades que a Carta atribui aos Estados-Membros é que estes designem um ou mais organismos que se encarregue(m) de coordenar a preservação do património digital, colocando à sua disposição o pessoal e os recursos necessários, fomentando a investigação e formação nesta área, impulsionando a troca de experiências e saberes entre instituições e profissionais, insistindo na colaboração dos fabricantes de equipamentos e programas informáticos, criadores, editores, produtores e distribuidores de documentos digitais

¹⁴ Segundo a *Carta para a Preservação do Património Digital da Unesco* (2003), o património digital consiste em recursos únicos fruto do saber ou da expressão dos seres humanos. Compreende recursos de carácter cultural, educativo, científico ou administrativo e informação técnica, jurídica, médica e de outro tipo, que é gerada directamente em formato digital ou a partir da conversão de material analógico já existente.

com as tradicionais bibliotecas e arquivos, com os repositórios digitais e com outras instituições que velem pelo património público.

Por seu lado, as instituições de memória tradicionais (bibliotecas, arquivos, museus, etc.), habituadas a assumir a responsabilidade de preservação, devem proporcionar a formação e especialização dos seus colaboradores, em virtude dos conhecimentos especializados e técnicos que a preservação digital exige. Além disso, devem redefinir responsabilidades, alocar recursos e estabelecer novas estruturas organizativas (Beagrie & Jones 2008), essenciais para assegurar a preservação digital como uma actividade contínua.

Os autores/criadores¹⁵ de informação digital – que, tradicionalmente, não estavam implicados neste processo –, têm agora um papel activo no domínio da preservação. Assim, os autores/criadores devem seguir normas e boas práticas, em consonância com o que foi estabelecido pelas instituições que levam a cabo essa missão¹⁶. Normalmente, o que se pede aos autores/criadores de informação digital é que utilizem formatos e suportes normalizados, suficientemente documentados, que tornem possível a gestão futura dos documentos.

Tal como os autores/criadores da informação digital, também os produtores de *hardware* e *software* são agora chamados a intervir no âmbito da preservação. É-lhes solicitado que colaborem com as instituições que levam a cabo essa missão, desbloqueando, por exemplo, os mecanismos tecnológicos de protecção para que possam ser aplicadas as estratégias de preservação, fornecendo a documentação necessária para que se conheça o contexto tecnológico em que os documentos foram criados e permitindo a criação de emuladores que recriem o ambiente tecnológico original.

Quem passa também a assumir um papel de destaque na preservação são os editores/distribuidores. No contexto analógico, eles não estavam incumbidos dessa missão; todavia, no contexto digital, assumem um papel de destaque, visto que as bibliotecas, em muitos casos, adquirem através de contratos de licenciamento apenas o acesso à informação e não o objecto que contém a informação. Ora, porque as bibliotecas só podem preservar aquilo que

¹⁵ Segundo Webb (2003, p. 78), o conceito de produtor é complexo porque inclui vários agentes com funções muito diferentes (designers, revisores, editores de programas informáticos, fornecedores de serviços,...). Além disso, inclui não só os autores, criadores e disseminadores de objectos digitais antes de estes serem entregues a um programa de preservação, como também aqueles que, através de programas de digitalização, geram novos documentos digitais.

¹⁶ Sobre o papel dos autores/criadores na preservação da informação digital, destaca-se o documento *Acid-free Bits : recommendations for Long-Lasting Electronic Literature* lançado pela Electronic Literature Organization (ELO), em 2004, da responsabilidade de Nick Monfort e Noah Wardrip-Fruin. O documento fornece informações sobre um conjunto de princípios básicos e de boas práticas que contribuem para a longevidade dos documentos digitais criados pela ELO.

possuem – e não estando estipulado nos contratos de licenciamento a cópia do documento para preservação – caberá aos editores a assunção dessa responsabilidade. Naturalmente, a existência de legislação que estendesse o Depósito Legal aos documentos digitais resolveria este problema. Numa altura em que as bibliotecas estão a perder o controlo físico das suas colecções (Muir, 2004, p. 73), é importante acautelar que todos os documentos sob custódia das diferentes editoras – e não apenas aqueles que revelem interesse económico – sejam efectivamente preservados:

Publishers themselves are unused to preserving content for the benefit of the public or of future generations of users, thus putting at risk large corpora of culturally and historically significant materials. (BRTF, 2010, p. 90)

Finalmente, mas também de suma importância, temos os consumidores/utilizadores da informação digital, que, atentos à vulnerabilidade da informação digital, devem ter a capacidade de adquiri-la e preservá-la indefinidamente. Para isso, e como adianta Lynch (apud Keefer & Gallart, 2007, p. 183), são precisas soluções que vão para além de “pagar para ver” e que assegurem o acesso permanente que permita aos consumidores possuir cópias dos documentos e não apenas o acesso a eles. Estas soluções permitir-lhes-iam ter um papel activo na preservação das suas colecções, às quais aplicando estratégias básicas (por exemplo, a execução de cópias de segurança, o refrescamento dos suportes ou a migração de ficheiros).

No processo de preservação do património digital, nenhum dos elementos que intervêm no ciclo de vida de um documento digital pode ficar de fora. Todos estão implicados nesta missão e todos são responsáveis pelo seu (in)sucesso, sendo, por isso, importante definir as responsabilidades e o papel de cada um neste processo:

The strongest incentives to preserve will be ineffective without explicit agreement on the roles and responsibilities of all the actors—those who create the information, those who own it, those who preserve it, and those who make it available for use. (BRTF, 2010, p. 2)

COOPERAÇÃO

Para que a preservação digital seja uma actividade economicamente sustentável, é necessário que haja cooperação entre os todos elementos implicados, quer pela partilha de recursos, conhecimentos, experiências e custos, quer pela divisão de diferentes responsabilidades, quer ainda pela eliminação de redundâncias, melhorando sinergias e evitando esforços inúteis:

It is economically impractical for every collecting institution to develop local digital preservation capabilities. A coordinated approach promises to be more cost effective, by spreading fixed costs over a greater number of institutions. (Lavoie & Dempsey, 2004)

Além disso, a união de várias instituições e organizações permite a partilha de influência em acordos com os autores/criadores da informação, com os produtores de *hardware* e de *software* e com os editores. Permite ainda atrair recursos e outros apoios para os programas de preservação levados a cabo em conjunto.

Todos os elementos são solicitados a colaborar na preservação digital: autores/criadores de informação, produtores de *hardware* e *software*, editores/distribuidores, instituições de memória (bibliotecas, arquivos, museus, etc.), utilizadores/consumidores e governos¹⁷.

1.2. Preservação digital: origens e definição do assunto

Segundo Fishbein (apud Thomaz, 2005, p. 11), o tema da preservação digital terá sido abordado pela primeira vez em 1964, em Paris, no *International Congress on Archives*. Mas foi na década de 70 do século passado que o assunto começou a ser tratado nos vários congressos e conferências internacionais de arquivistas. Por essa altura, enunciavam-se alguns dos desafios que os documentos digitais colocavam aos arquivos: alguns desses documentos estavam a ser eliminados devido à falta de conhecimentos sobre computação por parte dos arquivistas¹⁸, devido à escassez de recursos para a preservação ou devido aos obstáculos legais à recolha desse tipo de documentos. Outra dificuldade encontrada nesta altura era o reconhecimento arquivístico do documento digital. Nos anos 80, o tema da preservação digital começa a ter mais destaque e pela primeira vez, em 1988, o *International Congress on Archives* é dedicado inteiramente aos novos materiais arquivísticos. Este congresso foi marcado pela discussão sobre a definição, conservação e aplicação dos princípios e práticas arquivísticas aos documentos digitais.

Segundo Thomaz (2005, p. 14), os anos 90 do século XX foram os mais importantes para a investigação da gestão e preservação dos documentos digitais e caracterizam-se quer pela riqueza literária quer pela profundidade das discussões no âmbito da preservação digital. A segunda metade da década de 90 é marcada pelo início de vários projectos e programas de

¹⁷ A preservação do património digital deve assentar na colaboração e solidariedade internacionais, para que todos os países possam garantir a criação, difusão e preservação do seu património digital, assim como o seu acesso constante (UNESCO, 2003).

¹⁸ Bell (apud Thomaz, 2005, p 12-13) destacava como problema a relação entre arquivistas e informáticos, realçando o facto de os arquivistas considerarem o ambiente computacional demasiado técnico e hermético e de os informáticos terem dificuldade em reconhecer o valor arquivístico do documento digital. Bell defendeu a necessidade da inclusão de informáticos em arquivos e do aperfeiçoamento do conhecimento sobre arquivos e computação por parte de ambos os profissionais.

preservação digital dos quais resultam relatórios, manuais e recomendações que vieram trazer um grande contributo para a investigação nesta área. Estas iniciativas partem, sobretudo, de consórcios entre arquivos e bibliotecas universitárias e nacionais do mundo anglo-saxónico, estendendo-se à Europa e à Oceânia. Assumem principal destaque, nessa altura, os trabalhos de Jeff Rothenberg, Paul Conway, Margaret Hedstrom, John Garrett e Donald Waters, Brian Lavoie, entre outros. Ainda assim, e apesar de muitos dos aspectos da preservação digital terem começado a suscitar interesse em meados dos anos 90, segundo Marcum (2003, p. 644) muitos dos relatórios e artigos sobre o assunto não tecem mais do que comentários gerais sobre a sua complexidade, sobre o custo das tarefas implicadas e sobre a ambiguidade das responsabilidades. Apesar de as investigações levadas a cabo terem centrado esforços nos aspectos técnicos da preservação digital, elas revelam que os aspectos organizacionais, legais e económicos poderiam ser mais espinhosos do que os aspectos técnicos.

A primeira década do século XXI consolida as investigações levadas a cabo na década anterior¹⁹, havendo uma alteração de paradigma: a preservação digital deixa de estar centrada em acções imediatas, como a preservação dos suportes, para se concentrar em acções a longo prazo e em infra-estruturas técnicas e sociais que assegurem a perenidade dos documentos digitais. Por esta altura, assume principal destaque a investigação na área dos repositórios digitais e dos metadados de preservação. Assim, a investigação na área da preservação digital deixa de estar tão focada nas possíveis soluções técnicas para a resolução dos problemas levantados pela especificidade dos documentos digitais (e pelos suportes que os veiculam) para se concentrar em soluções que garantam a interoperabilidade e o acesso a sistemas emergentes como são os repositórios digitais. Estes sistemas de gestão e preservação da informação digital têm assumido um papel central no debate sobre a preservação digital, pois permitem armazenar esse tipo de informação, torná-la acessível, acompanhar as alterações feitas e manter a sua autenticidade:

In the sustainable digital environment, the same inclusive debate is occurring, and here the word is used to mean building an economically viable infrastructure, both social and technical, for maintaining valuable data without significant loss or degradation. This includes the whole socio-technical composition of the repository, the short- and long-term value of the material, the costs of undertaking an action, and the recognition that technologies do not sustain digital objects: institutions do, using the available technology. Clearly it is not possible to preserve digital information without a sustainable organizational, economic, social, structural, and technical infrastructure, nor is it sensible to preserve material without sustained value. (Bradley, 2007, p. 157).

¹⁹ A Unesco assume um papel fulcral no âmbito da preservação digital com a publicação, em 2003, de dois documentos: *Guidelines for the preservation of digital heritage* e *Charter on the Preservation of Digital Heritage*.

Corroborando a afirmação de Bradley, à luz das definições actuais, a preservação digital tem sido entendida não como uma actividade técnica isolada, mas sim como um conjunto de actividades, políticas e acções devidamente planeadas.

DEFINIÇÕES DE PRESERVAÇÃO DIGITAL

Enquanto instituições que assumem a responsabilidade colectiva de preservar e tornar acessível no futuro o património cultural, literário e científico, as bibliotecas devem compreender o que significa preservar em ambiente digital.

Na última década, várias definições de preservação digital têm sido apresentadas:

Digital preservation is used to describe the processes involved in maintaining information and other kinds of heritage that exist in a digital form (...). The purpose of preserving digital materials is to maintain accessibility: the ability to access their essential, authentic message or purpose.. (Webb, 2003, p. 20 e 34)

Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time. (ALA, 2007)

Digital Preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation (...) refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials may be records created during the day-to-day business of an organisation; "born-digital" materials created for a specific purpose (e.g. teaching resources); or the products of digitisation projects (Beagrie & Jones, 2008)

Consiste na capacidade de garantir a acessibilidade permanente com autenticidade suficiente da informação digital para que, no futuro, possa ser interpretada com base numa plataforma tecnológica diferente daquela que foi usada aquando da sua criação. (Faria & Pericão, 2008, p. 997)

A preservação digital consiste na capacidade de garantir que a informação digital permanece acessível, interpretável e autêntica na presença de uma plataforma tecnológica diferente daquela que fora inicialmente utilizada no momento da sua criação. (Ferreira, 2009, p. 13-14)

De entre as várias definições apresentadas, a mais completa parece ser a formulada em 2007 pelo *Working Group on Defining Digital Preservation* da ALA: conjunto de políticas,

estratégias e acções que asseguram o acesso aos documentos digitalizados ou nado-digitais²⁰, independentemente dos desafios colocados pela deterioração dos suportes e pela obsolescência tecnológica. O objectivo da preservação digital é, pois, a recuperação exacta de documentos digitais autênticos ao longo do tempo. Quando se refere a *políticas, estratégias e acções* esta definição torna explícita a necessidade declarada de uma intenção de preservar, de um plano para o fazer e de um compromisso com as actividades que permitam concretizar esse plano. Desta definição destaca-se também a necessidade de garantir o acesso, uma das principais missões das bibliotecas e um aspecto crucial na preservação de documentos digitais. Sendo o seu objectivo a *recuperação exacta de documentos digitais autênticos*, torna-se evidente o compromisso com a garantia de preservar os documentos assegurando a sua autoria e proveniência de forma autêntica e mantendo a integridade e funcionalidade do documento digital. Por fim, a completude desta definição está no facto de, naturalmente, a preservação digital estar direccionada não só para a preservação de documentos nado-digitais mas também para os documentos digitalizados.

Todas estas definições têm em comum o facto de enfatizarem a acessibilidade e autenticidade dos documentos digitais e o facto de serem necessárias plataformas tecnológicas diferentes daquelas utilizadas no momento da criação do documento original. Quer isto dizer que o objectivo da preservação digital não passa tanto pela preservação do suporte físico (como acontece com a preservação tradicional), mas sim por garantir que a informação nele contida permaneça acessível²¹ e autêntica ao longo do tempo.

OBJECTIVO(S) DA PRESERVAÇÃO DIGITAL

A análise das definições apresentadas mostra que o objectivo primordial da preservação digital é garantir que todo o património digital produzido pela Humanidade esteja acessível no futuro de forma autêntica. Segundo Caplan (2008, p. 7), os objectivos da preservação da informação digital são garantir a disponibilidade, a identidade, a compreensão, a fixidez, a viabilidade, a representação e a autenticidade daquele património.

²⁰ Documentos criados originalmente em ambiente digital e cujo manuseio efectivo só pode ser feito nesse ambiente. Distinguem-se dos documentos digitalizados que, por questões de acesso ou preservação, foram convertidos para um formato digital através de um qualquer processo de digitalização.

²¹ Segundo Caplan (2008, p. 7), a ênfase no acesso tem sido questionada como potencialmente perigosa: os editores parecem estar satisfeitos com o facto de arquivarem os seus conteúdos com o objectivo de preservação, mas não com o facto de eles estarem acessíveis de forma rotineira. Tal pode conduzir a novas definições em que se use preferencialmente o termo *usabilidade* em vez de *acessibilidade* da informação. Na verdade, nada pode ser utilizado se não puder ser acedido para uso; portanto, preservação e acesso continuam a caminhar lado a lado.



Figura 2: Pirâmide da preservação. Fonte: Caplan (2008, p. 9, adaptada)

A Pirâmide da preservação da Fig. 2 identifica os objectivos da preservação digital e os meios para os atingir. Assim, na base da pirâmide, a *disponibilidade*, que se apresenta como a condição básica para a preservação, pois não se pode preservar um documento digital que não se possui. Para isso, é necessário negociar acordos de depósito, licenças e contratos com editores, de modo a obter os documentos digitais pretendidos. A *identidade* refere-se à descrição dos documentos digitais através de metadados descritivos, os quais, em termos de preservação, permitem que um documento digital seja descrito com detalhe suficiente para ser acedido e usado no futuro. Quando se define como objectivo da preservação a *compreensão* pretende-se que um repositório garanta que a informação preservada possa ser compreendida de forma independente pela comunidade de utilizadores. Para tal, todas as actividades e processos de preservação devem ser criteriosamente documentados. O objectivo *fixidez* só é conseguido mediante o depósito e armazenamento seguros dos documentos digitais. Para o efeito, os sistemas de preservação devem protegê-los não apenas de alterações não autorizadas mas também da degradação física dos suportes que também pode estar na origem da corrupção dos dados gravados. A *viabilidade* diz respeito à qualidade que torna o documento digital legível a partir do suporte em que está armazenado. As principais ameaças à viabilidade são a deterioração e degradação dos suportes, bem como a obsolescência tecnológica; por essas razões, há que fazer uma gestão acurada dos suportes e das unidades de leitura que permitem interpretar o seu conteúdo. A *representação* está também relacionada com a obsolescência

tecnológica, mais propriamente com a obsolescência dos formatos em que a informação está codificada: o seu principal objectivo é fazer com que os ficheiros possam ser reproduzidos num novo ambiente tecnológico. Para tal, há que escolher a melhor estratégia que permita preservar o formato original ou converter o objecto num formato normalizado e, por isso, mais viável. Por último, a *autenticidade* significa que a integridade da fonte e do conteúdo do documento é verificada e mantida ao longo do tempo. Para conseguir este objectivo, o repositório deve garantir que a cadeia da custódia e que todas as alterações ao documento digital motivadas pelo processo de preservação sejam devidamente documentadas.

Os seis primeiros objectivos que formam a pirâmide da preservação visam, sobretudo, garantir que um objecto digital seja acessível para consulta no futuro, pois ele foi seleccionado e devidamente descrito e foram devidamente documentadas todas acções de preservação levadas a cabo com o intuito de mantê-lo legível, incorruptível e compatível com os novos ambientes tecnológicos. De nada vale a preservação do acesso se não se garantir que a informação contida num objecto digital é autêntica²². Só assim se pode concluir que qualquer estratégia de preservação foi levada a cabo com sucesso, ou seja, apenas e só quando as características originais de objecto digital se mantiverem através das acções de preservação.

1.3. Preservação digital e as bibliotecas

Preservar a memória colectiva, protegendo a documentação literária e científica para que as gerações futuras possam ter acesso a ela, é uma das missões das bibliotecas. A preservação digital deve, em primeiro lugar, ser entendida como uma actividade complementar de outros serviços prestados pelas bibliotecas, mas diferente da preservação tradicional, no sentido em que a gestão das colecções digitais assume um papel mais proeminente no quotidiano da gestão das colecções:

This suggests that as more and more digital materials come under the stewardship of collecting institutions, preservation will become less like an event occurring at discrete intervals, and more like a process, proceeding relatively continuously over time. As a consequence, it will become more difficult to distinguish preservation activities from the routine, day-to-day management of digital materials (Lavoie & Dempsey, 2004)

²² A questão da autenticidade no âmbito da preservação digital é delicada e tem suscitado um longo debate, pelo que será tratada em detalhe mais adiante em 2.4.1. *Autenticidade e integridade do objecto digital* com maior detalhe.

A especificidade dos documentos digitais implica acções e rotinas imediatas imprescindíveis ao processo de preservação digital. Para a maioria dos bibliotecários, os desafios associados à preservação digital são imensos: a incapacidade de determinar por onde começar a acção, a falta de conhecimentos especializados, a falta de ferramentas de confiança e simples de utilizar e a incapacidade de determinar custos (Gladney, 2007, p. 44). No entanto, o contexto digital não deve ser entendido pelas bibliotecas como uma ameaça, mas sim como uma oportunidade de redefinição e assunção como instituições imprescindíveis no seio da Sociedade de Informação. Assim, para que possam cumprir a sua missão básica de provisão do acesso à informação e de preservação dessa mesma informação, as bibliotecas devem fazer uso das novas tecnologias para uma maior e melhor disseminação da informação digital e devem identificar os desafios que o novo tipo de informação coloca aos processos de preservação.

1.3.1. Preservação tradicional vs Preservação digital

A preservação dos documentos digitais é mais problemática²³ do que dos documentos analógicos, isto porque o foco deixa de ser o suporte que contém a informação, passando a ser a própria informação:

Since time immemorial, the methods and practices of documentary heritage conservation have given the highest priority to preservation of carriers: paper and ink, the various generations of computer disks, magnetic tapes or emulsions for film, photography or microfilm. In the digital domain, it is the wine that is to be saved not the bottle. (Abid, 2007, p. 10-11)

²³ Gladney (2007, p. 7) afirma que o pessimismo relativamente à preservação digital se socorre, por vezes, da comparação da durabilidade do papel (e de outros suportes analógicos) com a dos suportes digitais. O autor combate esta falácia, solicitando que se reflita sobre o seguinte:

- Muitos dos papiros originais perderam-se. Não sabemos quantos, mas pode ter sido uma grande quantidade. A construção das pirâmides do Egipto, por exemplo, sugere que a sua concepção tenha sido feita através de sofisticados planos de engenharia que nunca vimos;
- O papiro que sobreviveu foi provavelmente aquele que apresentava melhor qualidade;
- O papel necessita de ser duradouro porque é muito difícil copiar o seu conteúdo para outro suporte. Uma vez que são facilmente copiáveis, os suportes digitais são, não obstante úteis, ainda que tenham uma durabilidade limitada;
- O conteúdo digital importante pode ser vertido em mais do que uma cópia de forma a assegurar a sua perenidade. Por acaso os egípcios criaram mais do que cinco cópias de papiros importantes?;
- Algum papel é duradouro, outro nem por isso: entre 1880 e 1930, os livros foram impressos em papel acidificado, que causou a sua rápida deterioração.

Ao invocar estes argumentos, Gladney recusa a asserção de que a informação digital seja mais frágil do que o papel ou o papiro: pelo contrário, os suportes digitais têm como grande vantagem a capacidade de armazenamento e são facilmente replicáveis.

A preservação da informação registada em suportes analógicos afigura-se tão simples quanto a preservação do objecto físico em si, uma vez que o conteúdo e o contentor são indissociáveis. Ainda que o conteúdo possa ser copiado para outro suporte (microfilme, fotocópia, etc.), o resultado será sempre uma cópia imperfeita do original: por isso os esforços de preservação centram-se no original. Apesar de a preservação da informação registada em suportes digitais poder também significar preservação do objecto físico em que se encontra registada (mas só a curto prazo!), existem outras ameaças além da deterioração física do suporte: a obsolescência tecnológica de todo o ambiente tecnológico que permite ler essa informação. O principal a ser preservado é o conteúdo, a informação em si: para tal, é necessário preservar o acesso.

Outro aspecto que distingue a preservação analógica da preservação digital é a frequência de acção: na primeira é, normalmente, esporádica; na segunda, deve ser continuada. Nas colecções analógicas, basta fornecer as condições ambientais ideais e esporadicamente levar a cabo tratamentos de preservação; nas colecções digitais, são necessárias intervenções contínuas e paralelas à evolução tecnológica (Keefer & Gallar, 2007, p. 35). Além disso, a preservação digital é um processo dinâmico, revestido de acções múltiplas, que devem ser levadas a cabo ao longo de todo o ciclo de vida do documento digital e que requerem o comprometimento e empenho de todos os que as executam:

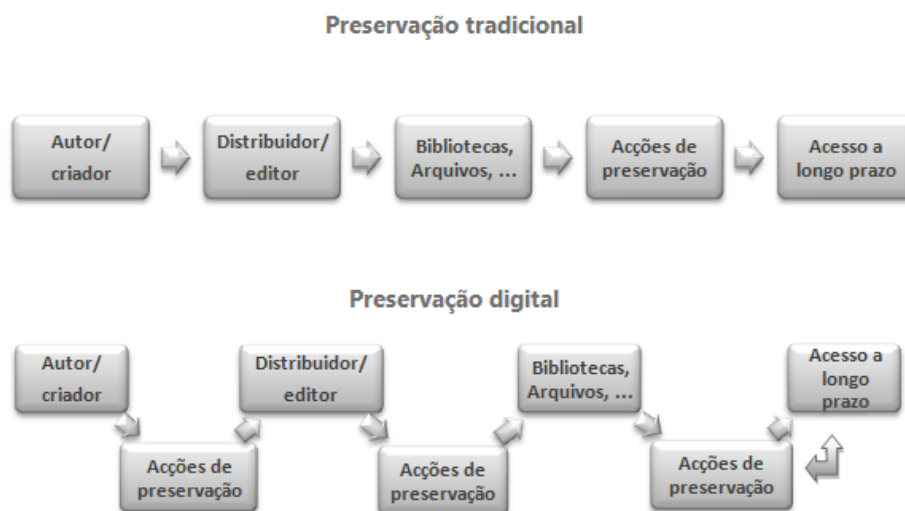


Figura 3: Ciclos de vida da preservação tradicional e da preservação digital. Fonte: BRTF (2010, p. 25 adaptado)

PRESERVAÇÃO RETROSPECTIVA E PRESERVAÇÃO PROSPECTIVA

Em ambiente digital, a percepção do tempo é muito relativa e distinta da que se tem em ambiente analógico. Tal deve-se não só à vulnerabilidade dos suportes mas também à rápida evolução tecnológica. O “tempo analógico” é distinto do “tempo digital”: por isso, são também diferentes os respectivos tempos de acção:

Mientras un documento tradicional puede estar décadas o incluso siglos en la biblioteca antes de que necesite un tratamiento específico, la vida de los soportes digitales dura unos pocos años (¡y a veces meses!) dependiendo de factores como el tipo de soporte, la calidad de su fabricación y las condiciones ambientales en que se ha almacenado desde su inicio.

Por otra parte, la evolución tecnológica deja ilegibles a muchos documentos a causa de la desaparición de su entorno informático original a pocos años de su creación, a diferencia del documento tradicional que no requiere ningún software o hardware para ser interpretado a lo largo de los años. (Keefer & Gallart, 2007, p. 31)

A preservação das colecções digitais não deve efectuar-se da mesma forma que a preservação das colecções analógicas. Objectos como livros, jornais, mapas, microfilmes e outros documentos em suporte analógico são passíveis de ser preservados por longos períodos de tempo (décadas, mesmo séculos), sem que os responsáveis pela sua preservação tenham de verificar a sua integridade periodicamente.

O que normalmente se faz com este tipo de colecções é uma *preservação retrospectiva*, ou seja, contrariam-se os efeitos do tempo, do uso e da degradação física dos documentos, recorrendo-se a tratamentos correctivos (físicos ou químicos) quando aqueles já se encontram deteriorados. É uma preservação *a posteriori* e, por isso, permite ao gestor das colecções gerir melhor as actividades de preservação (por exemplo, seleccionando documentos prioritários).

Este tipo de resposta à degradação dos suportes não se adequa às colecções digitais: nestas, uma vez detectado o dano, geralmente já é demasiado tarde para resgatar a informação registada. As características inerentes aos documentos digitais obrigam a que as acções de preservação sejam levadas a cabo no sentido de prevenir ou abrandar a sua deterioração. Assim, as colecções digitais exigem uma *preservação prospectiva*, que vai no sentido de prevenir e não no sentido de recuperar. As medidas de preservação são tomadas *a priori*, ou seja, evitando que os materiais se deteriore ou tornem obsoletos antes que a informação neles contida se perca irremediavelmente. Os esforços de preservação podem ser centrados na preservação do suporte onde está contida a informação digital, proporcionando um ambiente de armazenamento estável e controlado e de acordo com as normas de manuseamento correctas. No entanto, ainda que o suporte da informação digital seja preservado em boas condições físicas, existe sempre o risco de a tecnologia (*hardware* e *software*) que o torna legível ter ficado inoperacional, ou porque se tornou obsoleta ou porque foi descontinuada. A preservação digital exige, pois, medidas mais

complexas, como a aplicação de outro tipo de estratégias que contemplem mais do que a mera preservação do suporte²⁴.

A par da natureza distinta dos tempos de acção na preservação das colecções analógicas e na das colecções digitais, é interessante salientar que as primeiras são melhor preservadas quando os documentos são negligenciados, ou seja, quando são pouco usados. Por outro lado, nas colecções digitais o uso frequente dos documentos e a sua ampla distribuição aumentam as suas hipóteses de sobrevivência:

Letting documents lie on shelves in appropriate physical conditions was the best guarantee of preservation, and even allowing people to look at documents was long considered to be the worst enemy of conservation. On the contrary, the ability to allow digital information to circulate rapidly on new carriers, to migrate from one carrier to another, will be the ultimate guarantee of its permanent existence. (...) But the digital domain has one congenital defect: if you do not save something, you effectively erase it. (Abid, 2007, p. 10)

1.3.2. Digitalizar para preservar?

No final da década de 60 do século XX, começaram a ser desenvolvidos os primeiros programas de preservação das espécies bibliográficas ameaçadas, sobretudo daquelas publicadas nos últimos 150 anos em papel acidificado. A microfilmagem foi a técnica eleita devido, sobretudo, à durabilidade que as microformas ofereciam. A partir dos anos 80, a par dessa técnica de reprodução, emergia uma nova técnica – a digitalização – que, com a implementação da *World Wide Web*, nos anos 90, veio permitir uma maior difusão e um maior acesso aos documentos digitalizados.

Da digitalização para preservação resultam substitutos digitais que necessitam, também eles, de ser preservados (como, aliás, acontece com o microfilme). As bibliotecas têm recorrido à digitalização sobretudo para divulgar e tornar acessíveis as suas colecções raras, únicas e valiosas. Mas se o seu objectivo é também digitalizar para preservar (os originais ficam, desde logo, salvaguardados do uso e manuseio excessivos), então há que preservar as colecções digitalizadas, assegurando que os seus substitutos digitais sejam igualmente preservados:

If we do not succeed in establishing an effective long-term preservation strategy, these digital data [historical documents and other artifacts] may well become useless (or inaccessible) long before the destruction of their historical originals. (Börghoff et al., 2005, p. 39)

²⁴ Ver 3. *Estratégias de preservação digital*

A digitalização é uma grande fonte de documentos digitais: por isso, é preciso adoptar boas práticas que tenham em conta a preservação dos substitutos²⁵. Assim, deve ser considerada tão cedo quanto possível a forma como os substitutos vão ser preservados. Isso implica, por exemplo, determinar quais os suportes e formatos que devem ser utilizados; estabelecer o número e a qualidade das cópias realizadas (cópia-mestra e cópia de acesso); documentar todo o processo de digitalização, anexando-se os metadados necessários à futura preservação (equipamento utilizado, resolução, compressão, direitos de autor, etc.); definir a manutenção ou não dos originais.

1.3.3. Processo de selecção em preservação digital

Face ao valor social e à natureza efêmera do enorme volume de informação digital publicada, é fundamental uma selecção criteriosa e atempada dos documentos a preservar:

With traditional collections, lack of selection for preservation may not necessarily mean that the item will be lost, allowing for a comfort zone of potential changes in criteria for selection at a later stage. No such comfort zone exists in the digital environment where non-selection for preservation will almost certainly mean loss of the item, even if it is subsequently considered to be worthwhile. (Beagrie & Jones, 2008)

A selecção em preservação digital é um aspecto problemático, pois suscita questões muito delicadas, como o valor cultural ou científico dos documentos. Pode verificar-se que documentos preteridos no momento da selecção (e que posteriormente ganhem significado cultural) estejam irremediavelmente perdidos ou inacessíveis. Em preservação digital, o tempo de acção é essencial, dada a instabilidade dos suportes e a obsolescência tecnológica. É, por isso, impossível esperar para ter a certeza do valor permanente dos documentos e, então, tomar a decisão de preservá-los.

A *Web* é o maior meio de publicação e difusão de informação e constitui um espaço aberto onde todos podem participar. A selecção, para fins de preservação, da informação aí contida terá inevitavelmente de atender às novas manifestações culturais desse espaço virtual:

Heritage deals with things that were left to us from the past and preservation deals with these things often a long time after they were first created. But the digital world is moving so quickly that we cannot wait for 10 or 50 years to see what will prove to be heritage. (...) In the digital World, the time scale for preservation has shrunk. (Lusenet, 2002, p. 17)

²⁵ Informação complementar no capítulo 7 de *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives* da Unesco (2002).

De toda a informação digital produzida actualmente, apenas uma parte será preservada pelas instituições que assumem essa função. A selecção dos documentos digitais²⁶ a preservar deve constituir, pois, uma medida de gestão dessas instituições e deve basear-se na sua missão e objectivos.

Naturalmente a prioridade na selecção será dada aos documentos nado-digitais, pois os digitalizados (se o original não foi destruído) possuem um equivalente analógico mais estável e perene que será sempre o documento primário e original. A não selecção em ambiente digital de determinados documentos para preservação significará, quase de certeza, a sua perda irremediável. Ao seleccionador, coloca-se sempre o dilema da escolha: que informação vai ser mais útil e mais procurada no futuro? (Gladney, 2007, p. 47). Uma coisa é certa: não é viável, nem desejável preservar toda a informação digital criada (Hockx-Yu, 2006, p. 6):

Although not all resources can or need to be preserved forever, some will not need to be preserved at all, others will need to be preserved only for a defined period of time, and a relatively small sub-set will need to be preserved indefinitely. (Moghaddam, p. 82, 2007)

²⁶ Em *Digital preservation handbook*, Kenny et al. (2008) apresentam um instrumento útil (“Decision tree for selection materials for long term retention”) que permite apoiar as instituições na tomada de decisão aquando do processo de selecção.

2. Documentos digitais: especificidade e desafios

Um livro, feito de átomos, é algo material e tangível, ocupa um espaço físico e é passível de ser directamente lido por quem domine a linguagem em que está escrito. Pelo contrário, um documento digital, composto por *bits*, é algo intangível e, apesar de ocupar um espaço físico (seja num qualquer suporte magnético ou óptico, seja no disco de um computador), é virtual, apenas legível com recurso a um dispositivo que interprete a informação codificada em linguagem máquina.

2.1. Dependência e obsolescência tecnológica

Contrariamente aos documentos analógicos, os documentos digitais dependem dum ambiente tecnológico²⁷ específico que permite o acesso e a interpretação dos seus conteúdos. A informação digital é gerada por um *software*, que corre num dispositivo com determinadas características tecnológicas. Essa informação é registada e armazenada num disco duro, num CD-ROM, num DVD ou noutro suporte digital. Para lhe aceder, são necessárias unidades de leitura adequadas e *software* capaz de a interpretar e de a mostrar ao utilizador.

A dependência do ambiente tecnológico é, sem dúvida, a maior ameaça e a maior vulnerabilidade dos documentos digitais. Por um lado, as ferramentas (*software* intermediário entre a máquina e o humano) e os equipamentos informáticos sofrem rápidas evoluções que os tornam mais capazes e, ao mesmo tempo, mais expostos à obsolescência técnica e comercial. Por outro lado, os suportes são mais susceptíveis à deterioração e degradação física, quando comparados com os suportes analógicos.

²⁷ Para a criação de um documento digital é necessário recorrer a um ambiente tecnológico complexo e composto por componentes físicos (*hardware*) e por componentes lógicos (*software*). O primeiro inclui: o conjunto de dispositivos de entrada (teclado, rato, *scanner*, etc.), de saída (monitor, impressora, projector, altifalantes,...) e de armazenamento (disco duro, CD-ROM, DVD, *memory cards*, *pen memory*,...); a unidade de processamento central (CPU – *Central Processing Unit*); a unidade de processamento gráfico e as memórias primárias (RAM, ROM). O segundo corresponde ao conjunto de programas que, através de determinadas instruções, fazem funcionar os dispositivos físicos. De entre os vários tipos de *software*, distingue-se o software do sistema (responsável pelo funcionamento do próprio computador e de todo o *hardware* que inclui os sistemas operativos como o *Microsoft Windows* e o *Linux*, por exemplo) e o software de aplicação ou aplicações que permitem realizar uma série de tarefas: processamento de texto (*Word*, *Apple Works*, *Word Star*,...); gestão de bases de dados (*Access*, *Oracle*, *MySQL*, etc.); tratamento e edição de imagem (*Photoshop*, *Gimp*, *Paint.NET*, etc.); navegação na Internet (*Internet Explorer*, *Chrome*, *Firefox*, etc.); antivírus (*BitDefender*, *Norton*, *McAfee*, etc.); jogos, etc.

É por ter em conta esta especificidade dos documentos digitais que os esforços de preservação não se centram apenas na necessidade de manter o suporte de informação em boas condições físicas mas também na necessidade de manter o acesso, conservando ou não o contexto tecnológico. Conservar as ferramentas e os equipamentos não se afigura viável a longo prazo, pois corre-se o risco de multiplicação de autênticos museus tecnológicos. A solução poderá passar, portanto, pela transferência da informação para um suporte mais actual, que possa ser interpretado pelo novo ambiente tecnológico (com o risco de perdas durante o processo), ou pela criação de emuladores que recriem o ambiente tecnológico original (o que adiciona um novo intermediário e, desse modo, aumenta a complexidade da cadeia entre o documento digital e o utilizador).

OBSCOLESCÊNCIA DO HARDWARE E DOS SUPORTES

A obsolescência do *hardware* e dos suportes decorre da inovação tecnológica, que procura introduzir melhorias de velocidade, de eficiência e de capacidade de armazenamento, ao mesmo tempo que faz baixar os preços a que os equipamentos chegam ao consumidor. Um novo *hardware* proporciona inovações e melhorias no *software*, potenciando, por sua vez, a obsolescência do próprio *software* e dos formatos. Ao mesmo tempo, um novo *hardware* introduz um novo tipo de tecnologias, como novos periféricos e novos dispositivos de armazenamento (Kenney et al., 2003). Desta renovação constante dos componentes de *hardware* e suportes resulta que muitas bibliotecas possuam nas suas colecções documentos digitais que não podem ser acedidos porque os computadores e dispositivos de leitura necessários não foram conservados.

Degradação dos suportes

Tanto os suportes magnéticos quanto os suportes ópticos são ameaçados quer pela obsolescência tecnológica quer pela degradação física a que estão sujeitos (dadas as características físicas dos seus componentes)²⁸. A sua instabilidade intrínseca, aliada a factores externos (como o uso e condições de armazenamento), torna os suportes digitais mais vulneráveis e mais frágeis do que os suportes analógicos, pese embora a sua enorme capacidade de armazenamento.

²⁸ Sobre a obsolescência tecnológica e a degradação física dos suportes, Rossi (2008) afirma: de nada vale ter suportes de informação duradouros, se depois a esperança de vida técnica e comercial dos dispositivos que os interpretam é curta.

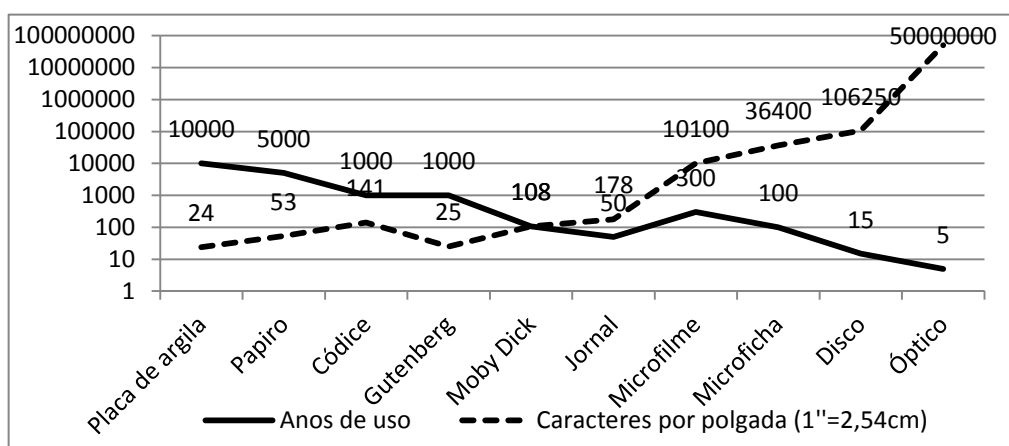


Figura 4: Densidade de informação vs Esperança de vida do suporte. Fonte: Conway (1996, adaptado)

Causas de degradação dos suportes

As principais ameaças à sua estabilidade dos suportes digitais são a humidade, a temperatura, a sujidade, a deformação mecânica, os campos magnéticos, o uso excessivo e o manuseamento incorrecto.

A humidade pode originar hidrólise e distensões nas fitas magnéticas e encoraja o aparecimento de fungos que consomem os materiais, impossibilitando a leitura da informação registada. As oscilações de temperatura provocam reacções químicas que podem degradar e tornar ilegíveis os suportes digitais. Por estas razões, as condições de armazenamento dos suportes devem respeitar os valores de humidade e temperatura ideais e assegurar que estes não sofram grandes oscilações²⁹.

A sujidade (pó, dedadas, fibras, etc.) acumulada nos suportes pode impedir a leitura correcta por parte do *hardware*. A perda de sinal é tanto maior quanto maior for a quantidade de sujidade acumulada no suporte. Alguns suportes são tão sensíveis que até partículas de fumo os tornam ilegíveis. Para evitar a acumulação de sujidade, os suportes devem ser acondicionados, transportados e manuseados correctamente.

Os materiais que compõem os suportes digitais são sensíveis a qualquer risco ou estrago, que podem impedir a sua interpretação correcta. A exposição a campos magnéticos estranhos, raio-X, microondas e radiações pode afectar os suportes magnéticos e os suportes óptico-magnéticos, impossibilitando a sua leitura ou mesmo destruindo-os. Nos suportes ópticos, é

²⁹ Para mais informação, ver 3.1.1. Boas práticas de manuseamento e armazenamento.

comum identificar e etiquetar os CDs ou DVDs utilizando canetas, impressora ou fitas adesivas. Os solventes, tintas e colas utilizados podem danificar os suportes, impedindo a sua leitura.

A par destas ameaças, outros factores contribuem para a degradação e erros de leitura dos suportes digitais: a qualidade dos materiais (nas bibliotecas, os constrangimentos financeiros impedem, muitas vezes, a aquisição de material de alta qualidade) e o mau estado dos equipamentos de gravação e de leitura.

OBSOLESCÊNCIA DO SOFTWARE

Os dados que constituem um documento digital estão codificados em linguagem máquina e só adquirem significado quando existe um *software* capaz de a interpretar. Os documentos digitais só existem se puderem ser recuperados, ou seja, se os dados codificados puderem ser interpretados por um determinado *software*. A sua inexistência torna os documentos digitais órfãos e, por isso, irrecuperáveis.

Novas versões de *software*, com novas capacidades capazes de satisfazer um público cada vez mais exigente, contribuem para a obsolescência do *software* e dos formatos. Apesar de muitos produtores de *software* proporcionarem a compatibilidade de versões, esta capacidade poderá não ser mantida para as versões mais antigas. Este potencial problema coloca-se a todos os que levam a cabo projectos de preservação digital.

De entre os vários tipos de *software*, distingue-se o *software* *proprietário* e o *software* *aberto* (*open source*). O primeiro nunca permite o acesso ao código-fonte e a sua utilização depende da aquisição de uma licença. A sua cópia, distribuição ou modificação apenas pode ser feita pelo produtor/distribuidor. O *Microsoft Windows*, o *RealPlayer*, o *Adobe Photoshop*, entre outros, são alguns exemplos de *software* proprietário. O *software* aberto distingue-se do anterior porque o código-fonte é disponibilizado, permitindo ao utilizador a liberdade de executar, copiar ou até melhorar o *software*. Além de assegurarem esta liberdade, as licenças de *software* aberto garantem que a sua cópia não infringe os direitos de autor associados. No contexto da preservação digital, a utilização de *software* aberto permite a aplicação de estratégias que vão desde a cópia livre do *software* até à sua emulação. A utilização de formatos decifráveis apenas por *software* proprietário obriga as bibliotecas a adquirirem licenças de utilização para que os documentos possam ser reproduzidos, mantendo íntegras as suas propriedades.

OBSOLESCÊNCIA DE FORMATOS

Segundo Gladney (2007), um formato é uma regra que especifica, para alguns tipos de dados, de que forma são escritas e lidas as sequências de *bits*. Na linguagem máquina, corresponde à sintaxe, ou seja, à forma como o conteúdo de ficheiro (os 0s e 1s) está estruturado. Sem esta sintaxe (formato ou especificação do formato), um ficheiro corresponde apenas a uma cadeia de 0s e 1s sem qualquer significado. O formato indica-nos a codificação, a sequência, a ordenação, o tamanho e as relações internas que o singularizam e que permitem que um ficheiro seja devidamente interpretado. Os formatos são identificados por extensões que identificam o tipo de conteúdo do ficheiro, tais como: *.rtf*, *.doc*, *.txt*, (texto); *.bmp*, *.gif*, *.jpg*, *.png*, (imagens); *.mp3*, *.wav* (som); *.avi*, *.mov*, *.mpeg*, (vídeo); ...

Para a obsolescência dos formatos contribuem as actualizações de *software* (novas versões deixam de suportar formatos antigos), a substituição de formatos por outros mais complexos e a pouca aceitação do público, que pode levar à sua estagnação e, consequentemente, à sua incompatibilidade com o ambiente tecnológico actual, ou até mesmo à sua descontinuidade (Kenney et al., 2003). Quando um formato se torna obsoleto, é necessário recorrer à documentação que contém as especificações para que o formato possa ser reinterpretado. A impossibilidade de aceder a tal documentação (ou a sua inexistência) pode inibir a sua recuperação.

2.2. Especificidade dos documentos digitais

A característica mais comumente apontada aos documentos digitais é a sua desmaterialização. Em ambiente *Web*, o suporte perde relevância e, deste modo, o acesso ao documento não fica condicionado por circunstâncias temporais ou espaciais. Esta volatilização do suporte não só o torna deslocalizado, mas também o torna facilmente transmissível e capaz de ser usado simultaneamente por diversos utilizadores. O volume de armazenamento reduzido, a pesquisa inerente, a maleabilidade dos conteúdos e a reprodutibilidade infinita e indistinta do original são também características únicas dos documentos digitais. Apesar de atractivas, estas características colocam sérios problemas ao seu processo de preservação. Em primeiro lugar, com a volatilização dos suportes, os esforços de preservação deixam de estar centrados no suporte propriamente dito para se centrarem no conteúdo, na informação em si. Enquanto a preservação dos suportes dos documentos analógicos significa preservar o suporte, a conservação do suporte nos documentos digitais não é garantia de que o conteúdo esteja preservado, pois é preciso assegurar que a informação contida nesse suporte se mantém acessível:

*In digital preservation there is a shift from preserving the carrier, the object itself and the information it contains, to preserving ways of accessing the information. We are moving from preservation **for** access to preservation **of** access. You cannot preserve what you see on a screen, that is only the temporary manifestation of a digital file, and we need to make sure we know how to recreate this temporary manifestation in new environments. (Lusenet, 2002, p. 22)*

A deslocalização espacial dos documentos digitais, por exemplo, pode dificultar a localização do país que tutela os direitos de autor do documento digital, o que, em termos de preservação, significa desconhecer que excepções legais estão reservadas às bibliotecas. Associada à deslocalização surge a instabilidade da *Web*, onde os documentos publicados estão sujeitos às decisões dos seus proprietários que podem apagá-los, alterá-los, mudar a sua localização ou alterar o seu nome (Webb, 2003, p. 30-31).

Outro aspecto que coloca desafios à preservação de documentos digitais é o facto de estes serem facilmente editáveis, o que lhes confere alguma plasticidade. Porque estão armazenados em formato digital, eles são facilmente revistos, reorganizados, reformatados, modificados e combinados com outros documentos. Esta maleabilidade revela-se problemática: quando existem várias versões de um mesmo documento (que versão preservar?); quando vários documentos estão hiperligados (onde começa e onde acaba um documento?); quando um documento possui vários elementos multimédia (que estratégia(s) de preservação aplicar a cada conteúdo?).

A plasticidade dos documentos digitais acarreta consigo aquele que será talvez o maior desafio colocado à preservação digital: a autenticidade. Como se poderá garantir que o documento preservado é autêntico se, por um lado, um documento digital é inerentemente instável e manipulável e se, por outro lado, a aplicação de algumas estratégias de preservação pode conduzir à perda de algumas propriedades significativas do documento?

A possibilidade de reproduzir infinitamente o documento digital, gerando cópias indistintas do original, coloca questões relativamente aos direitos de autor. Perante esta vulnerabilidade, muitos autores protegem os documentos digitais, colocando barreiras ao seu acesso (palavras-chave, pré-pagamento para acesso, encriptação, licenciamentos, etc.) e exigem que os seus direitos sejam reforçados de forma mais veemente. As bibliotecas só podem preservar aquilo a que têm acesso: com a inexistência de depósito legal aplicado aos documentos digitais, ficam interditas de preservar grande parte do património digital. Além disso, mesmo que as bibliotecas consigam integrar os documentos digitais nas suas colecções, a aplicação da maior parte das estratégias de preservação necessita de apoio legal que lhes permita iludirem os mecanismos de protecção tecnológica (em ambiente digital, preservar implica necessariamente copiar).

2.3. Estrutura de um objecto digital

Os dados processados e representados pelos computadores correspondem a sequências de 0s e 1s designadas por cadeias de bits (*bitstream*). Esta linguagem binária é a única que o computador consegue interpretar e é-lhe transmitida através de dois sinais eléctricos diferentes em que o 0 representa um valor e o 1 outro valor. Todas as memórias informáticas e todos os suportes de armazenamento de informação digital registam a informação através desta linguagem máquina.

Segundo Thibodeau (2002), um objecto digital é um objecto de informação, de qualquer tipo de conteúdo ou formato, que é expresso de uma forma digital. Ele pode ser gerado com recurso a um ambiente tecnológico: computador, câmara fotográfica, telemóvel, etc. Apesar de a definição ser aparentemente genérica, Ferreira (2009, p. 140) reconhece que objecto digital, enquanto objecto de informação, tanto pode conter informação criada em contexto tecnológico-digital (*objecto nado-digital*) como pode ser o resultado de um processo de digitalização a partir de suportes analógicos (*objecto digitalizado*). Desde que representado através de uma cadeia de bits (*bitstream*), qualquer texto, base de dados, imagem, vídeo, som, material gráfico, programa informático, página *Web* ou aplicação de *software* é considerado um objecto digital (Unesco, 2003).

Assim definido, qualquer objecto digital deve ser abordado em três dimensões: a dimensão física (inscrição de símbolos num suporte físico), a dimensão lógica (objecto reconhecido e processado pelo *software*) e a dimensão conceptual (objecto reconhecido e interpretado pelo ser humano). As estratégias de preservação digital devem ser aplicadas de modo a atender às ameaças que parecem sobre estas dimensões do objecto digital.

Enquanto *objecto físico*, o objecto digital corresponde à inscrição de símbolos num suporte físico³⁰, ou seja, constitui aquilo que o *hardware* consegue interpretar:

Basically, the physical level deals with physical files that are identified and managed by some storage system. The physical inscription is independent of the meaning of the inscribed bits. At the level of physical storage, the computer system does not know what the bits mean, that is, whether they comprise a natural language document, a photograph, or anything else. Physical inscription does not entail morphology, syntax, or semantics. (Thibodeau, 2002)

A forma como a informação é gravada num suporte digital varia de acordo com o tipo de suporte utilizado³¹, o que não impede que uma mesma imagem possa estar registada em suportes diferentes.

³⁰ Mesmo os objectos digitais que se encontram *online* têm de estar inscritos num qualquer suporte.

Do ponto de vista da preservação digital, as preocupações com a preservação física do objecto digital advêm do facto de os suportes digitais terem um tempo de vida limitado. Por essa razão, não é de estranhar que as primeiras estratégias de preservação da informação digital se concentrassem na preservação dos suportes onde a informação estava registada, acondicionando-os nas condições ambientais desejáveis e utilizando boas práticas de manuseamento. Quando os suportes corriam risco de degradação, adoptava-se a solução mais comum de copiar a informação digital para novos suportes³².

No entanto, cedo se verificou que a preservação do objecto digital não deveria limitar-se apenas à cópia dos dados registados para um suporte mais actual. Por si só, um suporte digital de informação não possui qualquer significado. Preservar o suporte não é suficiente para assegurar que a informação nele contida estará acessível no futuro. O suporte onde o documento está armazenado depende de todo o ambiente tecnológico que permita a sua leitura e decodificação.

Está-se, pois, na dimensão lógica do objecto digital. Enquanto *objecto lógico*, o objecto digital é entendido como um conjunto de unidades processáveis por qualquer *software*. Esse conjunto de dados encontra-se geralmente organizado segundo a arquitectura definida pelo *software* que foi utilizado na produção do objecto digital. Por sua vez, essas regras ou estruturas constituem aquilo que vulgarmente se designa por *formato* (Ferreira, 2009, p. 15).

The rules that govern the logical object are independent of how the data are written on a physical medium. Whereas, at the storage level, the bits are insignificant (i.e., their interpretation is not defined), at the logical level the grammar is independent of physical inscription. Once data are read into memory, the type of medium and the way the data were inscribed on the medium are of no consequence. The rules that apply at the logical level determine how information is encoded in bits and how different encodings are translated to other formats; notably, how the input stream is transformed into the system's memory and output for presentation. A logical object is a unit recognized by some application software. (Thibodeau, 2002)

A dimensão lógica de um objecto digital tende a perdurar no tempo mais do que a sua dimensão física, mas tal não significa que, preservando a sua dimensão lógica, o acesso ao objecto digital no futuro esteja garantido. As aplicações de *software* e os sistemas operativos

³¹ Os símbolos ou sinais físicos utilizados para representar a informação num CD-ROM diferem substancialmente dos utilizados para a representar num disco rígido. No primeiro exemplo, os símbolos utilizados são essencialmente pequenos orifícios reflectores dispostos em espiral sobre uma base de policarbonato. No segundo, são utilizados padrões magnéticos sobre um prato metálico. (Thibodeau, 2002; Ferreira, 2009, p. 15).

³² Ver 3.1.2. *Refrescamento*

utilizados para interpretar o objecto lógico, bem como os formatos em que este se encontra também podem tornar-se obsoletos, sendo necessário, por isso, modificar a codificação lógica original do objecto digital para que este possa ser interpretado por uma tecnologia diferente (Webb, 2003, p. 36).

Por lidar com unidades processáveis pelo computador, o objecto lógico ainda não é perceptível ao ser humano. Deste modo, o *software* assume a tarefa de preparar o objecto lógico de modo a que este tenha significado para o ser humano, transformando os símbolos manipulados no interior do computador em sinais analógicos que serão exibidos ao ser humano através de um periférico de saída (Ferreira, 2009, p.15)

Esta é a dimensão conceptual do objecto digital. Nesta dimensão, está-se perante aquilo com que se lida no mundo analógico, perante uma entidade significativa de informação – um texto, uma imagem ou um gráfico. As propriedades do *objecto conceptual* são, pois, aquelas que se revelam significativas no mundo real (Thibodeau, 2002).

É preciso não esquecer que o mesmo objecto conceptual pode ser representado em diferentes formatos lógicos e que cada um destes ser suportado por inúmeras representações físicas (Ferreira, 2006, p. 25), como se observa na Fig. 5:

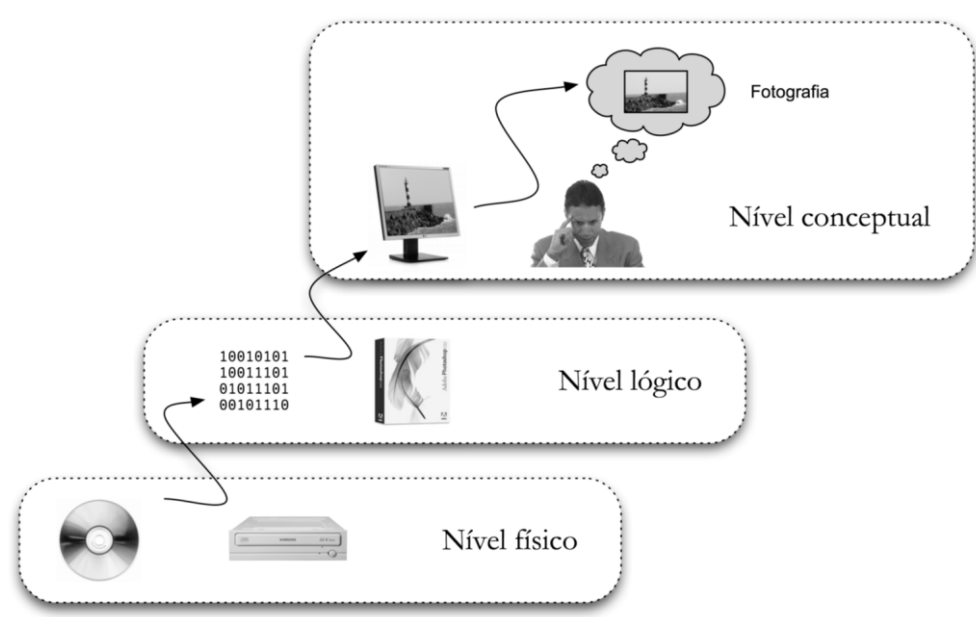


Figura 5: Cadeia de interpretação desde o nível físico ao nível conceptual. Fonte: Ferreira (2009, p. 16.)

Assim, uma fotografia – objecto conceptual – poderá ser codificada em formatos diversos (TIFF, JPG, PNG) – objecto lógico. Por sua vez, cada um desses formatos, ou todos eles, podem ser inscritos numa multiplicidade de suportes físicos (DVD, disco rígido ou *flash-drive*) – objecto físico.

Para que o ser humano descodifique um objecto digital através do conhecimento das suas três dimensões, é necessário que ocorra um conjunto de transformações. Numa primeira fase, o objecto físico é interpretado pelo *hardware*. Voltando ao exemplo da fotografia, ela pode estar codificada em diferentes suportes físicos. Sabe-se que os símbolos utilizados para representar essa fotografia num DVD podem diferir dos símbolos usados para a representar num disco rígido. Neste momento, cabe ao *hardware* transformar os símbolos inscritos no suporte físico num conjunto de dados que o *software* seja capaz de manipular. Já ao nível lógico, o objecto digital assume um formato (TIFF, JPG, PNG). Agora, cabe ao *software* preparar o objecto lógico para que este seja devidamente apresentado ao ser humano, transformando os símbolos digitais em símbolos analógicos que serão veiculados ao destinatário através de um periférico de saída. O objecto conceptual é a imagem que se forma na mente do ser humano (Ferreira, 2006, p. 22-23).

A preservação da informação digital não se esgota no conhecimento da dimensão do objecto digital que se pretende preservar. Sobre este aspecto, Thibodeau (2002) afirma:

In order to preserve a digital object, we must be able to identify and retrieve all its digital components. The digital components of an object are the logical and physical objects that are necessary to reconstitute the conceptual object. These components are not necessarily limited to the objects that contain the contents of a document. Digital components may contain data necessary for the structure or presentation of the conceptual object (...).

*In addition to identifying and retrieving the digital components, it is necessary to process them correctly. To access any digital document, stored bit sequences must be interpreted as logical objects and presented as conceptual objects. So digital preservation is not a simple process of preserving physical objects but one of preserving the ability to reproduce the objects. **The process of digital preservation, then, is inseparable from accessing the object.** You cannot prove that you have preserved the object until you have re-created it in some form that is appropriate for human use or for computer system applications.*

Actualmente, o interesse da preservação digital centra-se não tanto na preservação do objecto físico (nunca descurando o facto de os objectos digitais necessitarem de um suporte físico para subsistirem), mas sim na preservação do objecto conceptual, sendo esta a finalidade última da preservação digital (Webb, 2003, p. 36).

Perceber a anatomia de um objecto digital no contexto da preservação da informação digital é deveras importante, pois a escolha de qualquer estratégia de preservação implica saber que dimensão do objecto digital se pretende preservar. Se, por um lado, se pretende preservar a sua aparência e o seu formato de origem, é necessário preservá-lo na sua dimensão física e lógica; se, por outro lado, apenas interessa a informação em si contida no objecto digital, então é necessário preservá-lo apenas na sua dimensão conceptual.

É natural que no contexto da preservação digital, e tendo em conta a análise das dimensões do objecto digital, se questione a necessidade de preservar os componentes físicos e

lógicos de um objecto digital sem qualquer alteração, quando o que realmente interessa ao ser humano é o objecto conceptual. Thibodeau (2002) assegura que é possível alterar a forma como um objecto conceptual está codificado em um ou mais objecto(s) lógico(s) gravados em um ou mais objecto(s) físico(s) sem que isso tenha um impacto negativo na preservação do objecto digital:

Normally, we would say that changing something directly conflicts with preserving it. The possibility of preserving a digital object while changing its logical encoding or physical inscription appears paradoxical and is compounded by the fact that it may be beneficial or even necessary to make such changes. How can we determine what changes are permissible and what changes are most beneficial or necessary for preservation? Technology creates the possibilities for change, but it cannot determine what changes are permissible, beneficial, necessary, or harmful. To make such determinations, we have to consider the purpose of preservation. (Thibodeau, 2002)

Este é um paradoxo com que se deparam aqueles que levam a cabo um projecto de preservação da informação digital: preservar implica manter a integridade e a autenticidade do objecto preservado ao longo do tempo: no entanto, não é possível preservar um objecto digital a longo prazo sem alterá-lo nas suas dimensões física e/ou lógica. A aplicação das diferentes acções e estratégias de preservação digital implica, em muitos casos, a alteração dos componentes físicos e lógicos do objecto digital. Será que, ao alterarmos as propriedades significativas dos objectos digitais, os estamos a preservá-los, ou seja, estamos a mantê-los na sua forma íntegra e autêntica? Afinal, o que é importante preservar em contexto digital: o objecto físico ou o seu conteúdo intelectual? Será que quando preservamos um objecto digital devemos preservar toda a sua complexidade e variedade? Ou basta, em certos casos, preservar a informação que ele contém, a qual, como é sabido, pode ser codificada de várias formas, dependendo do formato utilizado? As respostas a estas questões não podem ser dadas de forma simplista e variam consoante a importância e o tipo de objecto digital a ser preservado. A compreensão de conceitos como *propriedades significativas* e a reformulação dos conceitos de *autenticidade* e *integridade* em ambiente digital são indispensáveis para determinar que dimensão(ões) do objecto digital podem e devem ser preservadas.

2.4. Propriedades significativas do objecto digital: o que preservar?

Em ambiente digital, entende-se por propriedades significativas aquelas características que um objecto digital deve manter ao longo do tempo, independentemente das acções e estratégias de preservação levadas a cabo. Em ambiente analógico, a preservação de um item não obriga à determinação prévia de quais as propriedades significativas desse objecto que devem ser mantidas ao longo do tempo. Tratando-se de objectos estáticos, cujo conteúdo intelectual se encontra encerrado num suporte, essas propriedades são as propriedades que o objecto encerra em si mesmo desde que se mantenha inalterado. Em ambiente digital, pelo contrário, o suporte e o conteúdo são entidades distintas e a aplicação de acções e estratégias de preservação implica, em muitos casos, libertar o conteúdo do suporte. Como resultado dessa acção, o objecto preservado deixa de ser o mesmo que foi depositado para preservação. Assim, e partindo do princípio de que não é possível preservar e simultaneamente manter inalterado e intacto um objecto digital, é necessário determinar o conjunto de propriedades significativas (aquelas que definem a essência do objecto) que esse objecto deve manter ao longo do tempo. Segundo Keefer & Gallart (2007, p. 71), o conjunto de elementos considerados essenciais a preservar vai para além do conteúdo intelectual do objecto digital e inclui outras características, como a sua funcionalidade e sua aparência, ou seja, aquelas qualidades que representam a essência do objecto tal como foi concebido pelo seu criador e tal como foi experimentado pelos utilizadores nesse momento.

O conjunto das propriedades significativas que um objecto digital deve conservar ao longo do tempo não é universal nem absoluto: a sua definição deve ter em conta a natureza da organização responsável pela preservação, as características da colecção e os requisitos e exigências da sua comunidade de interesse (Ferreira, 2009, p. 38). Além disso, a eleição desse conjunto de propriedades vai condicionar a escolha da estratégia de preservação digital a utilizar, pois nem todas as abordagens permitem preservar o objecto digital nas suas dimensões física e/ou lógica, mantendo, assim, a aparência e funcionalidades do original.

2.4.1. Autenticidade e integridade do objecto digital

A especificidade dos objectos digitais, bem como as operações que visam a sua preservação e o seu acesso, resultam inevitavelmente na alteração da sua forma física e/ou lógica o que pode, por sua vez, resultar numa ameaça à sua autenticidade e integridade. Sendo facilmente editáveis, os objectos digitais são também manipuláveis deliberada ou inadvertidamente (Ferreira, 2009, p. 37) e essas alterações nem sempre são detectáveis. Além

disso, as acções e estratégias de preservação supõem, em muitos casos, a transferência de um suporte e/ou formato para outro(s), a junção ou actualização de metadados, a alteração da funcionalidade e aparência do objecto digital.

A autenticidade e integridade já constituíam uma fonte de preocupações no âmbito da preservação tradicional³³. Em ambiente digital, o conteúdo dos documentos revela-se ainda mais vulnerável às ameaças acima referidas, pois as próprias acções e estratégias de preservação implicam alterações substanciais ao objecto digital original. Por tudo isto, as questões relativas à autenticidade e integridade ganham maior relevo quando abordadas no contexto da preservação da informação digital.

A forma como estes dois conceitos – autenticidade e integridade – são definidos e se relacionam não reúne consenso na literatura sobre preservação digital³⁴. Segundo Webb (2003, p. 108), a autenticidade de um objecto digital resulta da confiança depositada na sua *identidade* (é o que diz ser e não se confunde com nenhum outro) e na sua *integridade* (não sofreu nenhuma alteração que modificasse o seu significado). A *autenticidade* é definida por Webb como a garantia do carácter genuíno e fidedigno de certos materiais digitais, ou seja, a garantia de que são aquilo que dizem ser (quer se trate do objecto original, que seja cópia em conformidade e fiável do original, realizada através de processos bem documentados). A *integridade*, por sua vez refere-se ao estado dos objectos que se encontram completos e que não sofreram qualquer corrupção ou alteração sem que essas acções tivessem sido autorizadas ou documentadas (Webb, 2003, p. 157, 158). Assim, para que um documento se mantenha autêntico é necessário que a sua integridade também seja preservada. No entanto, em ambiente digital, preservar implica alterar o objecto digital:

We can generalize this insight and assert that the preservation of a digital information object does not necessarily entail maintaining all of its digital attributes. In fact, it is common to change digital attributes substantially to ensure that the essential attributes of an information object are preserved when the object is transmitted to different platforms (Thibodeau, 2002)

³³ Um dos exemplos mais apresentados na literatura quando se aborda a questão da autenticidade em ambiente analógico é o dos copistas. Do seu trabalho resultava, não raras vezes, um conjunto de diferenças relativamente ao original, motivadas por simples erros do copista ou devido à sua pretensão de fazerem “melhorias” no texto durante o processo de transcrição. Borghöff et al. (2005) não deixa de questionar: será a Bíblia autêntica?

³⁴ Alguns autores descrevem autenticidade e integridade como dois elementos relacionados entre si, mas distintos; outros estabelecem uma relação hierárquica entre eles; há ainda aqueles que consideram a autenticidade como parte componente da integridade; e outros que, ao contrário, consideram a integridade como um conceito dentro do conceito de autenticidade (Keefer & Gallart, 2007, p. 74-75).

Na impossibilidade de preservar ao longo do tempo o objecto digital tal qual foi produzido, mantendo-o inalterado, há que reformular e repensar o conceito de autenticidade e integridade. Assim, em contexto digital, a autenticidade não tem tanto a ver com a capacidade de demonstrar que o objecto preservado é o original, mas sim com a capacidade de demonstrar que ele *está conforme* o original (Ferreira, 2009, p. 37). Para Thidobea (2002), o resultado último da preservação digital deve ser preservar objectos digitais autênticos para que o *output* do processo de preservação sejam objectos *idênticos*, em todos os aspectos essenciais, àquilo que foi o objecto digital original (ou seja, o *input* no início do processo). Para este autor, um sistema de preservação da informação ideal deveria permitir que o documento D depositado num repositório digital num dado momento (T_0) fosse o mesmo a ser recuperado mais tarde (T_x), de tal forma que o resultado final fosse $D_{T_x} = D_{T_0}$. No entanto, a aplicação das diferentes estratégias de preservação implica que se mude o suporte físico e, muitas vezes, que se alterem os seus componentes lógicos. Deste modo, a equação para a preservação digital deve ser reformulada da seguinte forma: $D_{T_x} = D_{T_0} + \text{Delta}(X)$, em que Delta (X) é o resultado final das alterações (Delta) feitas ao longo do tempo (X).

Tanto a integridade de um objecto digital (ou seja, o objecto que está completo e íntegro), como a sua autenticidade (se o documento é realmente o que afirma ser) são difíceis de garantir em ambiente digital (Keefer & Gallart, 2007, p. 76). A imaterialidade do objecto digital resulta na inexistência de uma evidência física que possa comprovar que estas duas características se mantêm ao longo do tempo. Segundo Lynch (apud Thidobea, 2002), a autenticidade dos objectos digitais preservados é, afinal, uma questão de confiança³⁵. Os consumidores da informação contida nos objectos digitais preservados precisam de saber se essa informação é de confiança, ou seja, se essa informação é autêntica.

It is vital for every organization that its records be able to stand for the facts they are about, that is, that their content is trustworthy. It is equally important that, in time, those records can be proved to be what they purport to be, immune from any sort of tampering and corruption, that is, that they are trustworthy as records. (Duranti, 2001, p. 272)

Desse modo, só uma descrição pormenorizada, precisa e clara das características do objecto digital³⁶ pode determinar se ele está completo, se apresenta alguma alteração

³⁵ Segundo (Keefer & Gallart, 2007, p. 76), os utilizadores do futuro necessitam de confiar na instituição que custodia os recursos digitais, nos métodos utilizados durante a custódia e no recurso recuperado. Para tal, a solução proposta é a criação e certificação de repositórios digitais de confiança (*trusted digital repositories*). Ver 4. *Repositórios digitais e a preservação digital*

³⁶ Segundo Ferreira (2009, p. 37) “é fundamental documentar convenientemente a proveniência do objecto, contextualizar a sua criação e existência, descrever a sua história custodial e atestar que a sua integridade não foi comprometida, i.e., provar que o conjunto de propriedades que se consideram essenciais para a interpretação do objecto não foi adulterado ao longo do tempo.”

relativamente ao original e quais as alterações que foram levadas a cabo de forma a assegurar a sua preservação. Esta informação é dada através dos metadados, cuja importância é crucial no contexto da preservação digital³⁷.

³⁷ Ver Capítulo 4.2. *Metadados de preservação*

3. Estratégias de preservação digital

Desde os primórdios da civilização, a preservação da informação registada tem constituído um factor determinante para a protecção do património cultural. As técnicas de preservação utilizadas variaram ao longo dos tempos consoante o tipo de suporte em que a informação era inscrita. Em ambiente analógico, as estratégias de preservação implicavam o armazenamento dos documentos num ambiente adequado e seguro, deixando-os intocados durante décadas. Em ambiente digital, tal abordagem não é recomendada, pois, como foi mencionado anteriormente³⁸, a rápida obsolescência tecnológica e rápida degradação dos suportes exigem uma intervenção e estratégias activas e continuadas.

Não existe propriamente uma categorização das estratégias de preservação digital³⁹. Neste trabalho, elas aparecem organizadas de acordo com a dimensão do objecto digital preservada. Considerou-se também que nem todas as abordagens enunciadas na literatura correspondem a estratégias de preservação efectivas. Assim, aquilo que muitos autores consideram estratégias de preservação é aqui entendido como um conjunto de acções e boas práticas, que protegem o objecto digital apenas na sua dimensão física (por exemplo, o armazenamento e manuseamento adequados e o refrescamento ou a preservação da tecnologia). Por essa razão e porque só podem ser aplicadas a curto prazo, tais acções surgem em 3.1. *Acções que preservam do objecto digital na sua dimensão física.*

Ferreira (2009, p. xxvi) define estratégia de preservação digital como toda a abordagem técnica que garante o acesso continuado à informação existente em formatos digitais. Uma estratégia de preservação digital deve ter um impacto a médio/longo prazo e, por tratar-se de uma abordagem técnica (por vezes, complexa), normalmente requer conhecimentos especializados. Distingue-se, portanto, das acções de preservação que não requerem

³⁸ Ver 1.3.1. *Preservação tradicional vs Preservação digital*

³⁹ A análise da literatura revela que cada autor utiliza a sua própria categorização. Alguns autores preferem organizar as estratégias de preservação segundo a dimensão do objecto digital preservada. Assim, Thibodeau (2002) distingue **estratégias de preservação do objecto físico/lógico** (preservação da tecnologia e emulação) e **estratégias de preservação do objecto conceptual** (migração). Lee et al. (2002) referem-se a **estratégias que preservam o ambiente tecnológico** (preservação da tecnologia e emulação) e **as estratégias que visam superar a obsolescência de formatos** (encapsulação e migração). Outros autores preferem organizar a estratégias com base numa linha temporal. Webb (2003) considera as **estratégias a curto prazo** (preservação da tecnologia) e **as estratégias a médio e longo prazo** (migração, emulação). Beagrie & Jones (2008) distingue **as estratégias primárias** (migração e emulação) de médio/longo prazo das **estratégias secundárias** (preservação da tecnologia, encapsulação, arqueologia digital) de curto/médio prazo. Distinguem ainda aquilo que designam por **estratégias intermédias ou alternativas**, que incluem, por exemplo, a transferência para suportes analógicos.

conhecimentos técnicos especializados e que se aplicam apenas a curto prazo. Assim, considera-se, neste estudo, estratégias de preservação propriamente ditas a emulação e a migração que surgem em 3.2. *Estratégias que preservam o objecto digital na sua dimensão lógica* e 3.3. *Estratégias que preservam o objecto digital apenas na sua dimensão conceptual*, respectivamente. Tanto a emulação como a migração apresentam-se como as abordagens mais viáveis e as mais experimentadas actualmente no âmbito da preservação digital. Existem ainda aquelas abordagens que, por serem consideradas menos viáveis, menos experimentadas ou complementares das estratégias de preservação propriamente ditas, não foram incluídas em nenhum dos grupos anteriores. Elas ocorrem em 3.4. *Outras abordagens* onde se trata também de abordagens mais complexas e de último recurso como a arqueologia digital.

A aplicação destas abordagens técnicas (analisadas no seguimento deste capítulo), junto com uma série de boas práticas que visam a perenidade dos suportes em que a informação digital está registada, será a chave para o sucesso da preservação da informação digital a longo prazo. Outro factor importante para a preservação da informação digital com sucesso é a normalização de formatos, que simplifica a aplicação das estratégias de preservação.

NORMALIZAÇÃO DE FORMATOS

Os formatos dos ficheiros assumem um papel cimeiro no contexto da preservação da informação digital, na medida em que é o formato que indica a codificação, a sequência, a ordenação, o tamanho e as relações internas que o singularizam e que permitem que um ficheiro seja devidamente interpretado. A utilização de formatos normalizados, estáveis, largamente adoptados e, preferencialmente, não proprietários tem vindo a ser postulada como uma necessidade capital no âmbito da preservação digital⁴⁰. Por essa razão, a normalização de formatos é entendida como um processo transversal que facilitará a aplicação de todas as estratégias de preservação digital, e não como uma acção isolada que, por si só, consiga assegurar a continuidade dos documentos digitais.

Ao reduzir-se o número de formatos, reduzir-se-ão também os problemas da preservação da informação, visto que a uniformização do formato permitirá a aplicação transversal de uma estratégia de preservação a um maior número de objectos digitais, com a consequente redução de custos e da complexidade dos processos da preservação digital. A migração, por exemplo, pode ser facilitada pela conversão da informação digital num número

⁴⁰ Por exemplo, a maioria dos projectos de digitalização utilizam o formato .tiff (formato de ficheiros de imagens com etiquetas) porque se trata de uma norma aberta, estável e amplamente utilizada para a criação de imagens matriciais de preservação com melhores perspectivas de longevidade.

reduzido de formatos normalizados e independentes do *hardware* e *software* (Lee et al., 2002, p. 97). Além disso, sendo alguns formatos mais preserváveis do que outros, determinados programas de preservação e algumas instituições de depósito (como os repositórios digitais, por exemplo) garantem a gestão de um menor número de formatos ao definirem previamente aqueles em que os documentos devem ser depositados.

A normalização de formatos exige não só a conversão dos objectos digitais para o(s) formato(s) eleito(s) – o que pode motivar perda de alguns elementos essenciais – mas também o conhecimento das normas e o acompanhamento da sua evolução. Nem todos os formatos possuem normas estáveis, abertas ou documentadas e, mesmo que tal se verifique, também essas normas estão sujeitas a alterações inevitáveis à medida que surgem novas versões. Ferreira (2006, p. 39) indica alguns preceitos a seguir na escolha do formato:

Sempre que possível, deverão ser escolhidos formatos conhecidos pela comunidade de interesse e baseados em normas internacionais abertas. Isto poderá evitar futuras complicações ao nível de direitos de autor ou pagamento de *royalties*. Paralelamente, o formato de normalização deverá ser suficientemente rico para que as características dos vários formatos possam ser devidamente incorporadas. A aposta em sistemas abertos garante uma maior viabilidade da preservação futura e facilita a interoperabilidade de sistemas.

Conhecer os diferentes tipos de formatos utilizados para os diferentes tipos de conteúdos digitais, bem como as suas versões e variantes, é condição *sine qua non* para uma preservação digital eficaz a longo prazo. A selecção dos formatos pelas bibliotecas ou pelos repositórios digitais é feita em dois momentos: no momento da aceitação dos documentos ou no momento da sua digitalização. Quando os documentos têm uma primeira existência digital, a organização que os recebe para arquivo pode e deve definir os formatos em que estes se devem encontrar. O mesmo deve acontecer no momento da digitalização de documentos analógicos. Uma boa escolha de formatos para os diferentes tipos de conteúdos digitais tem profundas implicações na sua preservação a médio e longo prazo.

Consciente da importância dos formatos na preservação da informação digital, a *Library of Congress* (LC) desenvolveu um estudo alargado sobre este aspecto da preservação da informação digital e sobre os vários formatos disponíveis: *Sustainability of Digital Formats Planning for Library of Congress Collection*. O aspecto mais relevante deste trabalho da LC é a descrição exaustiva que se faz dos vários formatos, baseada em factores a ter em conta aquando da selecção e avaliação dos formatos para arquivo (apresentados em pormenor no Anexo 1). A LC distingue dois tipos de factores: factores de sustentabilidade e factores de qualidade e funcionalidade. Os primeiros prendem-se com a exequibilidade e o custo da preservação da informação digital e são importantes para a concretização de qualquer estratégia de preservação. Os segundos referem-se à capacidade de um dado formato representar as características específicas de um dado conteúdo.

A escolha de um formato deve basear-se no equilíbrio entre os factores de sustentabilidade, por um lado, e os factores de qualidade e funcionalidade, por outro lado. Por vezes, estes factores competem entre si, tornando difícil a selecção, até porque não existe nenhuma receita para a escolha do formato ideal.

Actualmente, existe uma série de formatos que são mais utilizados, mas isso não permite determinar quais serão os formatos mais usados no futuro. No entanto, analisando as características dos formatos actuais, é possível estimar quais são os que têm maior probabilidade de preservação para o futuro (FCCN, 2011).

Assim, o Arquivo da Web Portuguesa (AWP)⁴¹ recomenda a utilização, sempre que possível, de formatos adequados para preservação ou que se publiquem versões alternativas dos conteúdos que usem esses formatos. Para o AWP, um formato adequado para preservação é: (i) isento de direitos legais que restrinjam a sua utilização; (ii) uma norma emitida por um organismo oficial (exemplo W3C); (iii) documentado abertamente através de uma especificação pública e livre; (iv) amplamente usado; (v) lido e escrito por múltiplas plataformas de *software*, incluindo código aberto; (vi) não comprimido ou comprimido sem perdas de informação.

Pelo contrário, um formato com fracas características de preservação é: (i) proprietário e de especificação fechada; (ii) pouco usado; (iii) lido e escrito através de poucas plataformas de *software* de código fechado; (iv) comprimido, tendo havido perda de informação durante o processo de compressão; (v) composto por elementos embebidos (por exemplo, *macros*).

No Anexo 2, é apresentada uma lista proposta pelo AWP, que facilita a escolha de formatos de acordo com as suas potencialidades de preservação para o futuro.

⁴¹ Disponível na WWW: <http://arquivo-web.fccn.pt/>

3.1. Acções que preservam o objecto digital na sua dimensão física

A única forma de preservar os objectos digitais nas suas dimensões física, lógica e conceptual consiste em preservar os suportes e o ambiente tecnológico original em que os objectos digitais foram criados. Prevê-se que uma estratégia de preservação digital tenha uma aplicabilidade a médio/longo prazo de forma a afiançar que o recurso digital está efectivamente preservado. Deste modo, todas as acções e abordagens que apenas resultam numa preservação a curto prazo não devem ser consideradas como estratégias de preservação digital propriamente ditas, mas sim como um conjunto de boas práticas ou formas de actuar, que apenas se aplicam como meio de combater a curto prazo a deterioração dos suportes e a obsolescência tecnológica.

3.1.1. Boas práticas de manuseamento e armazenamento

O manuseamento e armazenamento adequados dos suportes digitais contribuem para a sua longevidade, ainda que a curto prazo. Por esta razão, aquelas duas acções não são entendidas como estratégias de preservação propriamente ditas, mas sim como boas práticas⁴², que, em conjunto com outras acções, correspondem a uma preservação preventiva, essencial para a aplicação das estratégias de preservação efectivas.

No Anexo 3, são enunciadas algumas recomendações relativas ao manuseamento e armazenamento dos suportes digitais presentes em *Directrizes da IFLA para a conservação e o manuseamento de documentos de biblioteca* (2004a) e em *Discos ópticos: cuidados e manuseamento*, de Fred R. Byers (2009).

A par de um manuseamento e armazenamento adequados dos suportes digitais, existem outras medidas que se revelam imprescindíveis no contexto da preservação digital: (i) a utilização de suportes duradouros e (ii) a preparação e manutenção de cópias de segurança (*backups*). A eleição de suportes duradouros para o armazenamento da informação digital contribui para a redução da necessidade de refrescamento⁴³, ajudando, assim, a diminuir perdas de informação associadas à deterioração/degradação dos suportes provocadas pelo manuseamento e acondicionamento incorrectos. Contudo, a utilização de suportes mais resistentes não protege de outras ameaças como os desastres naturais, ataques ou obsolescência tecnológica. Além disso, pode representar uma potencial ameaça à informação aí contida, pois

⁴² Designadas por Webb (2003) de estratégias de investimento ou estratégias a curto prazo.

⁴³ Ver 3.2.1. *Refrescamento (refreshing)*

provoca uma falsa sensação de segurança (Kenny et al., 2003). Perante os constrangimentos económicos com que as bibliotecas frequentemente se deparam, nem sempre é possível adquirir suportes de qualidade superior.

A execução e manutenção de cópias de segurança dos objectos digitais consistem na execução de cópias exactas do objecto digital que devem ser armazenadas remotamente para que o original e a cópia não estejam sujeitos às mesmas ameaças/desastres. Trata-se de uma componente transversal a todas as acções e estratégias de preservação, mas que, por si só, não pode ser entendido como uma estratégia. Na execução e manutenção de cópias de segurança dos objectos digitais apenas se lida com a perda de dados devido a falhas nos suportes ou no *hardware* (resultantes do manuseamento e/ou armazenamento incorrectos, por exemplo) ou devido a factores externos, como catástrofes, ataques, ou fenómenos semelhantes.

3.1.2. Refrescamento (*refreshing*)

O refrescamento consiste na transferência sistemática da informação digital registada e armazenada num dado suporte digital (disquete, CD-ROM, DVD, disco duro,...) para outro do mesmo tipo, de forma a combater a obsolescência e/ou a deterioração física dos suportes. Esta transferência permite que a informação se mantenha acessível independentemente da acção inibidora desses factores.

Alguma literatura (Kenney et al., 2003 ; Börgghoff et al., 2005) distingue dois tipos de refrescamento: o primeiro consiste na deslocação da informação digital de um suporte para outro do mesmo tipo, sem alteração de um único elemento da *bitstream*, permitindo que o objecto digital corra como antigamente no *hardware* e *software* de origem (*refreshing digital media*); o segundo consiste na deslocação da informação digital de um suporte digital para outro que não é igual ao original (*modified refreshing* ou *migration to other digital media*). Quando se transfere a informação de um suporte digital para outro diferente do original, há que ter em conta que a forma como é feita a gravação da informação difere de suporte para suporte⁴⁴, o que pode causar alterações na estrutura lógica do objecto transferido.

Keefer e Gallart (2007, p. 82) simplificam a abordagem do refrescamento, considerando que a transferência de dados de um suporte para outro (do mesmo tipo ou não) se efectua sem causar qualquer alteração no formato ou no *software*. Ou seja, apenas se refresca o ambiente, renovando-se o contentor, sem que se altere o conteúdo. As autoras distinguem as acções de

⁴⁴ A forma como a informação é registada num suporte magnético é diferente da forma como é registada num suporte óptico

refrescamento das cópias de segurança (*backups*), pois, apesar de acções parecidas, trata-se de acções distintas. O refrescamento está pensado para fins de preservação e as cópias de segurança estão pensadas para a recuperação de dados (no caso de perda acidental ou não). Obviamente, as cópias de segurança também terão de ser sempre preservadas, com recurso ao refrescamento ou a qualquer outra estratégia de preservação.

Foi afirmado que o refrescamento não constitui exactamente uma estratégia de preservação, pois não resolve a questão da obsolescência do *hardware* e *software* – apenas resolve o problema da deterioração/degradação física dos suportes (*refreshing digital media*) ou, em última análise, o problema da obsolescência dos suportes (*modified refreshing* ou *migration to other digital media*). Trata-se, por conseguinte, de uma acção preventiva e vital no contexto da preservação digital, mas que, por si só, não pode ser entendida como uma opção a eleger. Em termos de preservação digital, preservar o suporte não basta. É preciso garantir que a informação pode ser recuperada e processada no futuro através de todo o ambiente tecnológico que inclui o *hardware* e os dispositivos de leitura desse suporte, bem como o *software* que o interpreta (Lee et al., 2002, p. 94) – ou, no limite, através de um novo ambiente tecnológico.

O refrescamento é, pois, um pré-requisito para o sucesso de qualquer estratégia de preservação (Ferreira, 2006, p. 33), uma etapa, repetível e imprescindível do processo de preservação. Para ser levada a cabo, requer a determinação do tempo de vida de cada suporte e o estabelecimento de períodos mínimos de refrescamento sem que se corra o risco de perdas de informação. A frequência dos refrescamentos depende da política institucional e do tipo de suporte; todavia, deve realizar-se sempre dentro do prazo recomendado pelo fabricante. A degradação dos suportes pode ocorrer a um ritmo acelerado se as condições ambientais e o manuseamento não forem adequados ou correctos (Keefer & Gallart, 2007, p. 82).

A grande vantagem do refrescamento é tratar-se de um processo que não implica um grande investimento em equipamentos, nem um elevado grau de conhecimentos técnicos, a menos que a informação contida no suporte se encontre, de alguma forma, protegida. Por essa razão, o refrescamento exige a colaboração activa entre as bibliotecas e os produtores da informação. Estes últimos devem remover os mecanismos de protecção utilizados, para que as bibliotecas possam levar a cabo a sua missão de preservação⁴⁵.

Por sua vez, a desvantagem mais apontada reside no facto de o refrescamento não poder ser utilizado como estratégia de preservação, pois não resolve os problemas de obsolescência tecnológica ou de compatibilidade de formatos, por exemplo. Apenas resolve os problemas de degradação e de obsolescência dos suportes. Outra desvantagem é que requer uma gestão activa

⁴⁵ Para mais informação, ver 5.3. *Direitos de autor e a preservação digital: o caso português*

que implica a programação de acções de refrescamento de acordo com as indicações dos fabricantes e o estado dos materiais; a descrição das acções realizadas através de metadados técnicos; a avaliação da evolução do mercado para eleição dos suportes que oferecem maior durabilidade e estabilidade, entre outros (Keefer & Gallart, 2007, p. 82). Em suma, o refrescamento implica um maior investimento em recursos humanos que levem a cabo essas acções.

3.1.3. Preservação da tecnologia (*tecnology preservation*)

A maior ameaça à continuidade da informação digital é o desaparecimento dos meios de acesso (Webb, 2003, p. 34). Uma forma de combater esta ameaça é preservar o ambiente tecnológico original utilizado para aceder aos documentos digitais nele concebidos. Conservar e manter operacionais as plataformas que, entretanto, se tornaram obsoletas, de forma a garantir que os documentos digitais são acedidos no seu ambiente tecnológico original, implica a criação de verdadeiros museus⁴⁶ de *software* e *hardware* obsoletos. Embora preserve os documentos digitais nas suas dimensões física, lógica e conceptual, a preservação da tecnologia é inviável a longo prazo não só pelos custos que implica mas também pela sua impraticabilidade em termos espaciais e técnicos: a preservação da tecnologia requer espaço físico e manutenção das máquinas que, inevitavelmente, se degradarão com o tempo. Por estas razões, a implementação desta abordagem não deve ser levada a cabo por uma biblioteca isolada. Ela deve resultar da união de esforços entre várias instituições. Todavia, esta solução limitará a portabilidade e acessibilidade dos documentos digitais, dependentes sempre dos ambientes tecnológicos armazenados em locais específicos do globo (Lee et al., 2002, p. 95).

Os defensores desta abordagem argumentam que é condição *sine qua non* para a preservação das funcionalidades, da aparência e da autenticidade do documento digital a manutenção e a disponibilização das conduções tecnológicas originais (Lee et al., 2002). Além do alargamento do período de tempo em que se viabilizem a criação e a implementação de outras estratégias, outra vantagem da preservação da tecnologia reside no facto de ela poder ser a estratégia interina mais prática para a preservação do acesso a recursos digitais mais complexos, como jogos. (Webb, 2003; Børghoff et al., 2005; Beagrie & Jones, 2008). Na verdade, esta pode ser a solução para alguns objectos digitais, mas terá de ser sempre encarada como uma medida transitória (Thomaz e Soares, 2004). É impossível manter os equipamentos, programas, manuais, documentação, metadados e licenças por muito tempo, devido à

⁴⁶ Caplan (2008, p. 10) e Kenney et al. (2003) designam esta acção por *computer museum*.

disponibilidade cada vez menor de peças, à necessidade de conhecimentos cada vez mais especializados e à descontinuidade de algumas tecnologias. Com uma gestão activa, o intervalo de tempo de acesso fornecido pela preservação da tecnologia pode ser apenas de 5 ou 10 anos, a partir do momento em que o formato original perde vigência (Webb, 2003, p. 131). Esta actuação é, pois, uma medida de curto/médio prazo, pois o suporte técnico acabará por desaparecer rapidamente e o acesso tornar-se-á cada vez mais problemático com o passar do tempo (Beagrie & Jones, 2008).

3.2. Estratégias que preservam o objecto digital na sua dimensão lógica: *emulação*

A emulação é uma estratégia de preservação da informação digital que surge como opção à preservação da tecnologia, na medida em que permite o acesso aos documentos digitais no seu ambiente tecnológico original sem necessidade de preservar ou manter em funcionamento o *hardware* e/ou *software* originais. O ambiente tecnológico original é mantido através da criação de um novo *software* (emulador⁴⁷) que imita o funcionamento do *hardware* e/ou *software* originais⁴⁸ os quais, entretanto, se tornaram obsoletos, reproduzindo fielmente as suas funções. Deste modo, o objecto digital é preservado na sua dimensão lógica, sem necessidade de preservar a sua dimensão física. O principal objectivo da emulação é manter a aparência e a funcionalidade do objecto digital, ao permitir a sua leitura no ambiente tecnológico original emulado (Lee et al., 2002, p. 95).

Rothenberg, o proponente e principal entusiasta da emulação como estratégia de preservação digital, argumenta que esta é a única estratégia viável a longo prazo, uma vez que é

⁴⁷ Um emulador é, segundo Ferreira (2009, xx, p. v), um *software* capaz de reproduzir o comportamento de uma plataforma de *hardware* e/ou *software* numa outra, que de outro modo seria incompatível. Assim, um emulador de M a correr em N é um programa E que é executado em N e reproduz o efeito de cada uma das instruções básicas de M através de instruções próprias sequenciadas em N, isto é, N interpreta as instruções de M. A combinação da máquina N com o emulador E comporta-se tal qual a máquina original M. Assim, os programas inalterados de M podem ser executados na combinação (N+E) (Börghoff et al., 2005, p. 62).

⁴⁸ Distinguimos dois tipos de emuladores: emuladores de sistemas operativos e/ou *software* e emuladores de *hardware*. Os primeiros focam-se na reprodução do sistema operativo por completo permitindo, deste modo, a execução de diversas aplicações no contexto de um único emulador (...). Os segundos visam imitar o comportamento de uma plataforma de *hardware*, possibilitando que vários sistemas operativos e correspondentes aplicações possam ser executados no contexto de um único emulador (...). (Ferreira, 2006, p. 34)

a única que mantém as propriedades significativas⁴⁹ do documento original, mantendo, assim, a sua autenticidade. Börghoff et al. (2005, p. 57), também entusiastas da emulação como estratégia de preservação, adiantam que o método mais seguro de preservar os documentos digitais, mantendo a sua autenticidade e tornando-os acessíveis às gerações futuras, é armazená-los no seu formato original e garantir que no futuro esteja disponível um emulador do ambiente tecnológico original.

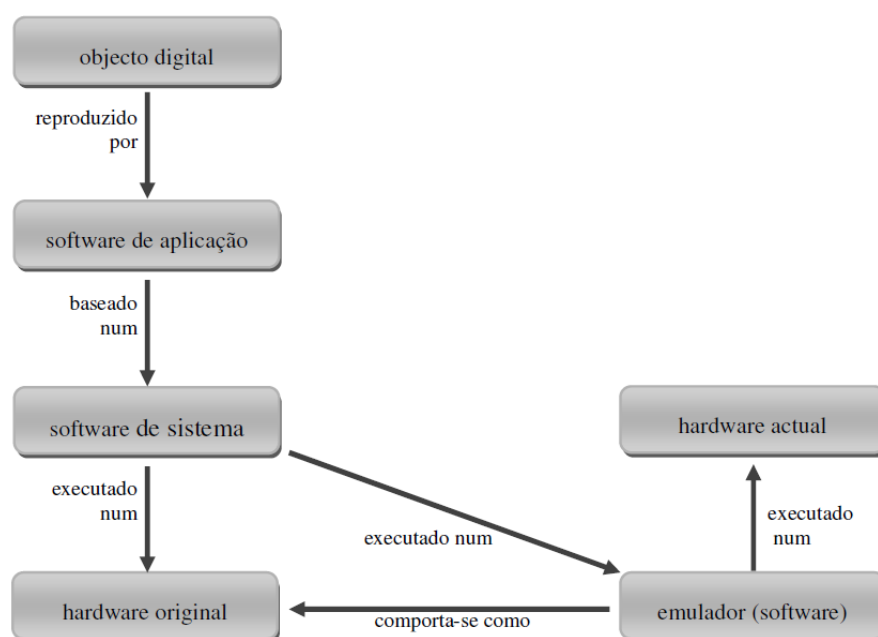


Figura 6: Recuperação de um objecto digital através de um emulador

A grande vantagem da emulação é que este processo dispensa a preservação da tecnologia sem que, por isso, se percam as características originais do objecto digital e a sua funcionalidade. Deste modo, esta estratégia apresenta vantagens relativamente à preservação da tecnologia, na medida em que as técnicas de emulação se centram apenas na preservação do objecto digital na sua dimensão lógica, mantendo o seu formato original (Ferreira, 2006, p. 34). Além disso, porque recria o ambiente tecnológico original, a emulação destina-se à preservação de todo o tipo de documentos, apresentando-se como a estratégia mais viável para os recursos digitais mais complexos, como os documentos compostos e/ou interactivos (Beagrie & Jones, 2008). Outra vantagem da emulação prende-se com o facto de esta estratégia de preservação não requerer um seguimento contínuo do formato em que se encontra o documento digital, pois a actualização de formatos é um problema que não se coloca, desde que a plataforma original seja

⁴⁹ Ver 2.4. *Propriedades significativas do objecto digital: o que preservar?*

mantida. Assim, os esforços de preservação não dependem do número de documentos, mas apenas do número das fontes suportadas e das plataformas-alvo (Börghoff et al., 2005, p. 7)

Para que a emulação seja bem sucedida, há que criar um pacote de informação⁵⁰ que contenha dados relativos ao objecto digital emulado, às especificações do *software* usado na sua emulação (sistema operativo, aplicações e outras informações relevantes) e às especificações sobre o emulador a usar no futuro. Bia e Sánchez (2002, p. 5) alertam para o facto de a perda de algum destes componentes informacionais poder inviabilizar o acesso à informação. O código do emulador deve ser escrito em código aberto, através de uma linguagem de programação bem documentada e normalizada e com boas hipóteses de continuar a ser utilizada no futuro (Webb 2003, p. 141).

Além de necessitar duma documentação rigorosa e completa sobre os sistemas emulados, a emulação está dependente de mão-de-obra altamente especializada e, consequentemente, dispendiosa e inacessível à maioria das bibliotecas. A emulação de ambientes tecnológicos em risco de obsolescência pode dificultar a recuperação dos documentos digitais, pois pressupõe que, no futuro, os utilizadores⁵¹ operem adequadamente aplicações informáticas arcaicas que funcionam por emulação (Webb, 2003, p. 141). Outra desvantagem desta estratégia prende-se com o facto de os emuladores, como *software* que são, também serem susceptíveis de sofrer obsolescência – por isso, exigem também eles esforços de preservação (transportando-nos para um infinito processo recursivo):

Emulators themselves become obsolete; therefore, it becomes necessary either to replace the old emulator with a new one or to create a new emulator that allows the old emulator to work on new platforms. In fact, if you get into an emulation strategy, you have bought into a migration strategy. Either strategy adds complexity over time. (...) Any preservation strategy that relies on specific software is carrying all the problems associated with those products into the future. (Thidobea, 2002)

A concepção de emuladores de *software* e *hardware* proprietários implica a negociação e/ou obtenção dos direitos de propriedade intelectual e licenças afectas a essas tecnologias (Kenney et al., 2003; Thidobea, 2002, Borghoff et al., 2005). Esse pode ser um processo moroso, dispendioso ou mesmo impossível. Outro aspecto que pode condicionar no futuro o acesso a documentos digitais através de emuladores é a possibilidade de as aplicações conterem vírus, o que pode resultar em perdas significativas de informação.

⁵⁰ Para mais informação, ver 3.4.1. *Encapsulação*

⁵¹ Para pessoas não especializadas e sem conhecimento contextual, o acesso pode estar condicionado e poderá, no futuro, apenas ser feito por arqueólogos digitais (Börghoff, 2005, p. 78).

3.3. Estratégias que preservam o objecto digital apenas na sua dimensão conceptual: *migração*

A migração da informação digital consiste na transferência periódica de material digital de uma dada configuração de *hardware* e *software* para outra, ou de uma geração de tecnologia obsoleta para outra subsequente (Ferreira, 2009, p. xxvi; Lee et al., 2002, p. 96; Kenney et al. 2003). Estes últimos autores referem-se à migração não como um acto isolado no tempo e no espaço, mas como um conjunto de tarefas organizadas e periódicas, concebidas com o objectivo de manter o acesso aos documentos digitais através de ambientes tecnológicos actuais. É uma medida contra a obsolescência dos formatos e suportes⁵² e visa manter o acesso à informação, conservando as suas propriedades significativas.

Ao contrário da preservação da tecnologia ou da emulação, a migração não visa a preservação do ambiente tecnológico original e, portanto, não preserva o objecto digital nas suas dimensões física e/ou lógica. Esta estratégia, focalizada sobretudo na preservação da dimensão conceptual do objecto digital, tem como objectivo, segundo Ferreira (2006, p. 36), manter os objectos digitais compatíveis com as tecnologias actuais, para que um utilizador comum seja capaz de recuperá-los, aceder e utilizar a informação independentemente dos avanços tecnológicos e sem necessidade de recorrer a artefactos menos convencionais (por exemplo, os emuladores).

⁵² Segundo Lazinger (2001), a migração distingue-se do refrescamento porque envolve a transferência de todo o ambiente tecnológico (formato, software, ...) e não apenas a renovação do suporte. Apesar de o refrescamento fazer parte da maioria das estratégias de migração, este processo resolve apenas o problema da deterioração dos suportes, mas não resolve o problema da obsolescência do *hardware* e do *software*.

Segundo Kenney et al. (2003), enquanto alguns autores utilizam indiferentemente os termos refrescamento e migração, outros defendem que, teoricamente, a migração vai para além da conversão de dados e visa contrariar a obsolescência não apenas dos suportes, mas também do código e do formato dos dados. Assim, a migração inclui o refrescamento, mas difere dele na medida em que nem sempre é possível fazer uma cópia exacta ou uma réplica dos dados de um objecto digital quando convertidos para outro suporte ou configuração de *hardware/software* e continuar a manter a compatibilidade do objecto digital com a nova geração tecnológica.

Webb (2003, p. 134) afirma que a migração se distingue do refrescamento (que apenas mantém o fluxo de dados, transferindo-os de um suporte para outro), porque pressupõe a transformação da forma lógica do objecto digital de tal modo que a sua dimensão conceptual possa ser restituída ou apresentada através dos novos equipamentos e programas informáticos.

Duranti (2002, p. 273), ao abordar a questão da autenticidade, distingue-os nestes termos: o refrescamento gera uma reprodução completa do conteúdo e elementos formais dos documentos, sendo que as cópias resultantes cópias fiéis do original. A migração, pelo contrário, reproduz o conteúdo de um documento digital, provocando algumas alterações na configuração e no formato e afectando, muitas vezes, outros componentes do documento; em alguns casos, pode haver perda de informação.

A migração é a estratégia de preservação da informação digital mais usada pelas instituições detentoras de grandes fundos bibliográficos. Compreende-se a opção: os procedimentos são relativamente simples, estão bem definidos e, segundo Keefer & Gallart (2007, p. 84-85), trata-se de um processo informático muito experimentado, que não requer, portanto, conhecimentos técnicos muito especializados. A adopção desta estratégia por parte de muitas instituições prende-se também com o facto de ela estar direccionada para grandes lotes de recursos que se encontrem num só formato, como acontece, por exemplo, nos repositórios digitais⁵³. Assim, para que a migração seja bem-sucedida, sugere-se a adopção de formatos normalizados e preferencialmente não proprietários, para os quais os documentos digitais serão convertidos no momento de armazenamento, o que facilitará e tornará possível automatizar os processos de migração⁵⁴ e manter os custos mais baixos. Este processo permitirá também programar antecipadamente as migrações, realizando-as no momento oportuno. Para que isso aconteça, é necessário que as instituições que adoptem esta estratégia estejam atentas aos avanços tecnológicos e à evolução/actualização dos formatos, a fim de tornarem as decisões mais acertadas no momento da selecção do novo ambiente. Ao contrário do refrescamento, cujo calendário responde às recomendações dos fabricantes e tem em conta a esperança de vida dos suportes, a migração não pode ser programada tão sistematicamente. Pelo contrário, a evolução dos formatos responde a factores externos, impulsionados pela indústria da informática e segundo as pressões e oportunidades do mercado (Keefer & Gallart, 2007, p. 84-85).

A grande vantagem da migração assenta na sua capacidade de manter os documentos “vivos”: com efeito, a migração sucessiva e contínua dos objectos digitais para novos formatos exige a sua permanente revisão, a fim de se averiguar se a migração foi feita com os padrões de

⁵³ Mais informação em 4. *Repositórios digitais e a preservação digital*

⁵⁴ Sobre a automatização de processos de preservação baseados em migração de formatos, consultar a tese de doutoramento de Ferreira (2009), para cujo projecto de investigação foi desenvolvida uma Arquitectura Orientada ao Serviço capaz de auxiliar organizações e/ou indivíduos na implementação de intervenções de preservação. O sistema desenvolvido é constituído por um conjunto de componentes, fisicamente distribuídos, que são capazes de realizar o seguinte conjunto de actividades: executar acções de preservação baseadas em migração de formatos (conversão); determinar a quantidade de informação, propriedades significativas e funcionalidades perdidas durante uma migração (controlo de qualidade); produzir relatórios que possam ser utilizados como metainformação de preservação e que documentam a intervenção de preservação (autenticidade); e fornecer sugestões de formatos de destino e/ou serviços de conversão que maximizem a satisfação da entidade-cliente (selecção de alternativas de migração).

Os principais contributos desta investigação são: a capacidade de preservar objectos digitais recorrendo a conversão de formatos sem que haja necessidade de implementar complexos sistemas de migração; a capacidade de obter relatórios detalhados sobre as migrações realizadas permitindo, assim, documentar todo o processo de preservação e deste modo assegurar a autenticidade dos materiais; e a possibilidade de comparar diferentes alternativas de migração e identificar de forma objectiva qual é a mais adequada para satisfazer as necessidades de uma organização.

qualidade definidos e se o conteúdo essencial não foi colocado em risco ou ofendido durante a transferência (Börghoff et al., 2005, p. 54). Assim, todo o processo de transferência deve ser submetido a um controlo de qualidade⁵⁵, que consistirá na comparação dos objectos digitais originais⁵⁶ com os objectos migrados, de forma a verificar se estes conservam as propriedades significativas dos originais.

A adopção da migração como estratégia de preservação digital simplifica o processo de recuperação dos documentos digitais, uma vez que estes se encontram sempre em versões actualizadas e, assim, podem ser recuperados nos ambientes tecnológicos actuais. Tal significa que os documentos estão sempre disponíveis (porque não dependem de tecnologias obsoletas ou emuladas) e que os utilizadores não necessitam de conhecimentos especializados para efectuar a sua recuperação.

A transferência de formatos obsoletos para novos formatos pode, ainda, resultar na melhoria da qualidade dos dados, pois as novas versões dos formatos ou os novos formatos podem, por exemplo, remover imperfeições existentes nos formatos nativos ou conferir novas funcionalidades que os formatos originais não permitiam. O que aqui é apresentado como uma vantagem pode resultar numa desvantagem na medida em que, ao melhorar a qualidade dos dados ou ao permitir novas funcionalidades, a migração da informação para nova versão de formatos ou um novo formato pode colocar problemas quanto à sua autenticidade e à sua integridade.

Ao privilegiar a preservação do objecto digital na sua dimensão conceptual, esta estratégia preserva o seu conteúdo, mas pode não preservar a sua apresentação, a sua funcionalidade ou o seu contexto. O processo de migração de informação encerra em si este tipo de desvantagem, pois, na migração dos objectos digitais para um formato mais recente, é muito provável que algumas das suas propriedades não sejam correctamente transferidas e, em alguns casos, possa mesmo haver perda de dados. Na origem destes inconvenientes estão as incompatibilidades existentes entre os formatos de origem e destino ou a utilização de

⁵⁵ Ferreira (2009, p. 52) sublinha que, “após uma conversão é fundamental avaliar os resultados obtidos, i.e., verificar em que medida os objectos que resultaram da conversão satisfazem os requisitos definidos *a priori* pela entidade preservadora. Este processo consiste, usualmente, na análise e comparação dos objectos que resultaram da migração com os objectos originais, tendo por base o conjunto de propriedades significativas definido previamente pela entidade preservadora (Hofman, 2002b; Rusbridge, 2003). Este conjunto de propriedades significativas constitui o nível de compromisso assumido pela organização no que toca à preservação dos materiais digitais. Após a conversão, uma avaliação abaixo das expectativas poderá implicar a selecção de uma nova alternativa de migração e a repetição de todo o processo de conversão (Ferreira et al., 2006a).”

⁵⁶ Börghoff et al. (2005, p. 49) defendem que apenas devem ser preservadas as versões originais dos documentos importantes para assegurar que, caso a versão resultante se perca, deteriore ou torne inacessível, possa sempre ser recuperada através, por exemplo, da arqueologia digital.

conversores incapazes de realizar as suas tarefas adequadamente (Ferreira, 2006, p. 36). A corrupção dos dados originais, a possível alteração da aparência e a perda de propriedades significativas resultantes de migrações sucessivas podem acumular-se e causar grandes alterações ou perdas de informação⁵⁷, podendo mesmo ofender a integridade e autenticidade do objecto digital. Por essa razão, todas as acções levadas a cabo durante o processo de migração, bem como qualquer efeito que estas tenham produzido no documento, devem ser documentadas através de metadados técnicos. Assim, os futuros utilizadores saberão quais as diferenças entre o documento original e a versão migrada (Keefer & Gallart, 2007, p. 87-88). Outra medida exigida ao processo de migração é o controlo de qualidade (ou auditoria), que permite averiguar se o processo de conversão foi bem sucedido, ou seja, permite verificar se as propriedades significativas do objecto digital previamente definidas foram mantidas.

É neste contexto que surge o conceito de canonização, isto é, a técnica que permite averiguar se as propriedades significativas de um documento se mantêm intactas quando se converte um objecto digital dum formato para outro, comparando o objecto convertido com o objecto original. Lynch (apud Ferreira, 2009, p. 39) apresenta-a como uma forma de avaliar o sucesso de uma migração. Além disso, esta técnica permite que o processo de migração seja levado a cabo sem que haja uma perda significativa de dados:

(...) although migration inevitably brings some changes it is an acceptable solution to the challenges of technological obsolescence, provided it retains what are variously called the 'significant properties', 'salient features' or 'essential characteristics' of the objects to be migrated. The structure, appearance and functionality of objects can all be at risk during migration, but even if some differences are introduced – or so the argument runs – this will be inconsequential if migration is carried out in such a way that no 'significant' loss occurs and the features necessary 'for digital objects to remain accessible and meaningful are preserved intact. (Yeo, 2010, p. 87).

Lee et al. (2002, p. 97) assinalam outra desvantagem da migração da informação: nos casos em que nem sempre é fácil distinguir o que é formato do que é conteúdo (por exemplo, os documentos multimédia complexos), a migração não é uma opção pacífica. Componentes múltiplos podem requerer actividades de migração separadas, tornando-se o processo de migração arduo. Aliás, para alguns recursos multimédia, a migração pode nem sequer ser possível sem que se comprometa significativamente a funcionalidade desses recursos.

Este processo de transferência tem de ser levado a cabo periodicamente, durante todo tempo de vida dos documentos digitais, sempre que haja uma nova versão do formato ou sempre que se verifique a actualização de um *software*. Deste modo, a migração enquanto estratégia de preservação pode revelar-se dispendiosa e morosa, pois exige migrações

⁵⁷ Tal efeito pode ser contrariado pela utilização da migração *a-pedido* analisada em detalhe em 3.3.1. *Migração a-pedido*

sucessivas e constantes, rigorosos controlos de qualidade, além da necessidade de documentar-se rigorosa e detalhadamente todos os processos de transferência.

3.3.1. Migração *a-pedido*⁵⁸

Este tipo de migração consiste na conversão, não do objecto digital mais actual, mas sim do objecto digital original. Com a migração *a-pedido* pretende-se combater a degradação que os objectos preservados sofrem ao fim de algumas iterações sucessivas, o que pode traduzir-se numa mudança substancial dos objectos digitais originais. Como afirma Ferreira (2006, p. 40), sempre que é efectuada uma migração, os objectos digitais resultantes serão de alguma forma diferentes dos objectos de partida.

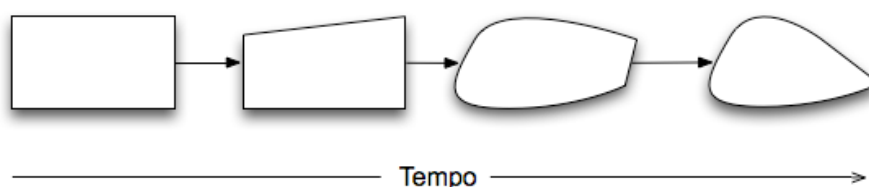


Figura 7: Degradação do objecto digital ao longo de sucessivas migrações. Fonte: Ferreira (2009, p. 30)

Com a migração *a-pedido*, o fenómeno de degradação do original pode ser atenuado: se de uma determinada conversão resultar um objecto substancialmente diferente do original, o problema poderá ser resolvido numa conversão futura, com recurso a um conversor de melhor qualidade ou a um formato de destino mais adequado, como demonstra a Figura 8:

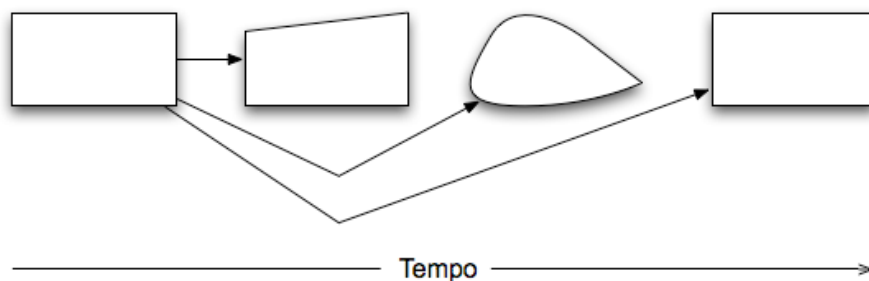


Figura 8: Migração *a-pedido*. Fonte Ferreira (2009, p. 40)

⁵⁸ Este tipo de migração é designado por Webb (2003, p. 138-139) como *viewers and migration at the point of access*. Lee et al. (2002) chamam-lhe *migration on request*.

O Projecto Cedars⁵⁹ e o Projecto CAMiLEON⁶⁰ introduziram este tipo de migração com o objectivo de evitar a repetição das sucessivas acções de migração. Em vez de migrações periódicas motivadas pelas alterações tecnológicas, a migração *a-pedido* ocorre no momento em que é requerida a informação (é nesse momento que se migra a informação de uma tecnologia obsoleta para outra actual). A ideia é que a informação seja armazenada enquanto *bitstream* e que, depois, seja convertida à medida que for necessária, utilizando-se ferramentas de migração que possam correr nas plataformas actuais. Este tipo de migração obriga à separação do *bitstream* do seu ambiente tecnológico e suportes originais no momento em que o objecto digital é depositado para armazenamento. É criada uma cápsula que reunirá essa cadeia de *bits* juntamente com os metadados técnicos que especificam como se deve migrar esse objecto noutros ambientes tecnológicos.

Quando as migrações incidem sempre sobre o objecto digital original, existem mais hipóteses de sucesso da migração, pois a qualidade dos conversores e a capacidade dos formatos vão evoluindo ao longo do tempo. Assim, uma migração mal sucedida pode ser superada através de uma migração futura do objecto digital original com recurso a novos conversores e a novos formatos.

Para que seja bem sucedida, a migração *a-pedido* obriga à (i) manutenção actualizada e completa da documentação sobre o formato dos ficheiros e sobre o método da sua transformação; e (ii) revisão dos metadados técnicos e das evoluções tecnológicas. Além disso, os metadados técnicos bem como os métodos de acesso aos conteúdos do objecto digital devem estar vinculados a este, mas armazenados à parte, para que possam ser actualizados de forma mais centralizada (Webb, 2003, p. 139)

A migração *a-pedido* é recomendada para a preservação de documentos digitais pouco solicitados. A sua interpretação e transformação ocorrem unicamente no momento em que é solicitado o acesso aos documentos, reduzindo-se, assim, a perda de propriedades significativas e o custo associado às migrações sucessivas.

⁵⁹ Disponível na WWW: <http://www.ukoln.ac.uk/services/elib/projects/cedars/>

⁶⁰ Disponível na WWW: <http://www2.si.umich.edu/CAMiLEON/>

3.3.2. Migração distribuída

Este tipo de migração surge dos mais recentes desenvolvimentos no domínio da migração, nomeadamente com a introdução de arquitecturas distribuídas de conversores (Ferreira, 2006, p. 41). Assim, a migração distribuída pressupõe a existência de um conjunto de serviços de conversão que se encontram acessíveis na Internet e que poderão ser invocados remotamente, recorrendo-se a uma aplicação-cliente (Ferreira, 2006, p. 41).

Várias iniciativas têm vindo a concorrer para o desenvolvimento deste tipo de conversores. O *Typed Objects Model* (TOM) é um sistema distribuído de conversores desenvolvido pela biblioteca da Universidade da Pensilvânia. Este modelo consiste num sistema distribuído de conversores e é utilizado na migração de objectos digitais para novos formatos. Outro sistema de migração distribuída é o *web service* desenvolvido pelo *Lister Hill National Center for Biomedical Communications*, que tem capacidade para converter objectos digitais de cinquenta formatos distintos para o formato PDF (Ferreira, 2006, p. 41).

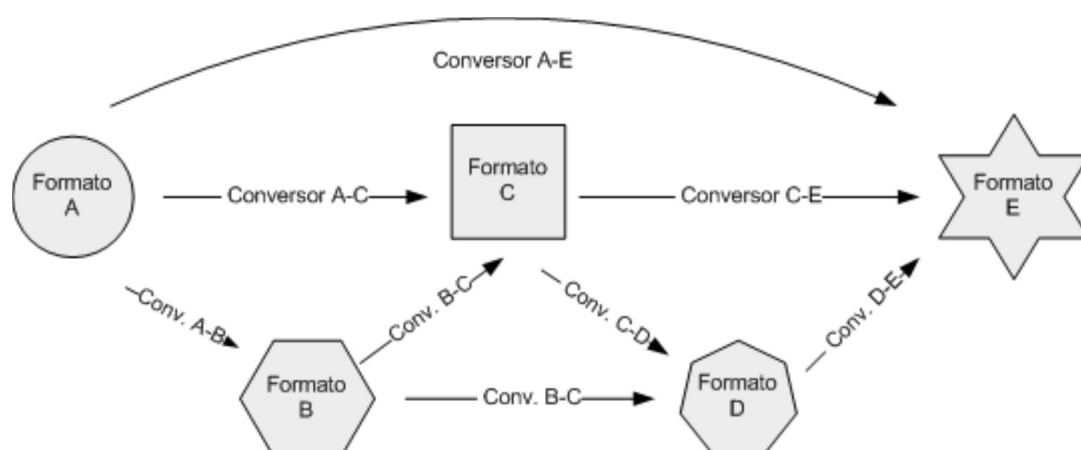


Figura 9: Migração distribuída baseada em Web Services. Fonte: Ferreira (2009, p. 31)

Em Portugal, a Universidade do Minho desenvolveu uma Arquitectura Orientada ao Serviço⁶¹ denominada *CRiB – Conversion and Recommendation of Digital Object Formats*⁶² que disponibiliza várias centenas de serviços de conversão, avaliação e recomendação:

⁶¹ *Service-oriented Architecture* (SOA)

⁶² Para informações mais detalhadas sobre este serviço, ver capítulo 4 da tese de doutoramento de Ferreira (2009) – *Preservação de longa duração informação digital no contexto de um arquivo histórico*.

O CRiB (...) tem como objectivo auxiliar tanto instituições, como utilizadores individuais, na implementação de estratégias de preservação baseadas em migração. O conjunto de serviços disponibilizados por esta plataforma permite a qualquer entidade cliente implementar de forma transversal e automática todas as tarefas subjacentes à preservação de objectos digitais. (Ferreira, 2009, p. 72)

A migração distribuída mostra-se vantajosa relativamente aos modelos de migração convencionais, porque, segundo Ferreira (2009, p. 31): (i) a utilização de serviços de conversão permite esconder as especificidades de cada conversor e da plataforma que o suporta; (ii) a criação de serviços redundantes assegura a fiabilidade do sistema perante situações de ruptura parcial; (iii) a existência de múltiplos caminhos de migração permite à solução resistir ao desaparecimento gradual de parte dos conversores; (iv) este tipo de abordagem é compatível com uma série de variantes de migração (por exemplo, normalização e migração *a-pedido*); (v) a criação de uma rede global de conversores poderá conduzir a uma redução generalizada dos custos de preservação (qualquer organização poderá rentabilizar os seus investimentos no desenvolvimento de conversores, publicando-os na rede de serviços e cobrando uma pequena taxa pela sua utilização).

Ainda assim, este tipo de migração pode não ser adequado a todos os contextos na medida em que:

Um repositório de informação digital pode facilmente conter milhares de itens, atingindo níveis de armazenamento na ordem dos múltiplos Terabytes. Transferir através da Internet um volume de informação desta natureza acarreta custos que poderão ser impeditivos para muitas organizações. (Ferreira, 2009, p. 32)

3.3.3. Actualização de versões ⁶³

A actualização da versão do formato é uma acção comumente utilizada pela generalidade dos utilizadores: consiste na actualização dos objectos digitais produzidos por um determinado *software*, recorrendo a uma versão mais actual do mesmo. Normalmente, os próprios produtores de *software* fornecem programas de conversão que permitem que as novas versões do seu produto leiam versões anteriores do mesmo, actualizando-as e gravando-as no novo formato (Thibodeau, 2002). A actualização de versões é uma estratégia que, a longo prazo, se mostra desvantajosa, uma vez que, com o devir dos anos, os formatos estarão sujeitos a descontinuidade e, como argumenta Thibodeau (2002),

⁶³ Na literatura, este tipo de migração tem várias designações: *Backwards compatibility* (Beagrie & Jones, 2008 e Webb, 2003); *Version migration* (Webb, 2003); ou *Updates within a product family* (Börghoff, 2005).

One problem with this approach is that using more recent versions of software, even with the original formats, may present the preserved documents with characteristics they did not, and perhaps could not, have had (...). Another potential problem is that over time, product lines, and the migration path, may be terminated.

A actualização de versões, ainda que esteja disponível, não dura para sempre, mais tarde ou mais cedo, se forem lançadas novas versões do mesmo *software*, os produtores acabam por considerar inconveniente e pouco lucrativa a manutenção das características que tornariam a versão actual compatível com a anterior (Lazinger, 2001, p. 80-81). Em virtude de os produtores de *software* não assegurarem ao longo do tempo a compatibilidade entre versões antigas e actuais, podendo mesmo abandonar uma dada versão, este tipo de migração, por si só, não funciona como estratégia de preservação (Börghoff, 2005, p. 48). Além disso, nem todos os produtores de *software* oferecem essa funcionalidade e nem sempre ela existe para todo o tipo de objectos digitais. Outra desvantagem da actualização de versões prende-se com os inconvenientes associados às migrações sucessivas, que podem originar alterações indesejadas nos objectos digitais, ocasionadas pelas actualizações constantes de gerações de programas.

Não obstante as desvantagens referidas, a actualização de versões simplifica a migração e reduz os custos associados, pois são os próprios produtores de *software* que assumem a responsabilidade de garantir a compatibilidade entre versões anteriores e a migração dos documentos para as versões actuais. Este factor pode prolongar o período de tempo até que seja necessário fazer uma transformação ou tratamento mais amplos dos documentos digitais (Webb, 2003, p. 133).

3.3.4. Conversão para formatos concorrentes

A conversão para formatos concorrentes ou paralelos surge como forma de garantir que os objectos digitais sobrevivem ao tipo de rupturas tecnológicas referidas no ponto anterior, como seja a descontinuidade dos formatos. Esta estratégia consiste na conversão dos objectos digitais para formatos de uma linha de produtos concorrentes (Ferreira, 2002, p. 38), preferencialmente normalizados, bem documentados e não proprietários.

Em alguns casos o formato sucessor pode ser totalmente diferente do formato original o que pode causar perdas de informação substanciais (Börghoff et al., 2005, p. 48). A conversão para formatos concorrentes pode implicar perdas ou até acréscimos de informação, resultando na acumulação de erros e conduzindo a resultados cada vez menos parecidos com o original (Caplan, 2008, p. 10). A estratégia contrária consiste em salvar o original e no desenvolvimento

de programas que permitam migrar directamente do original para o formato actual (migração *a-pedido*).

3.4. Outras abordagens

Embora muitos autores as considerem estratégias de preservação, estas abordagens, neste trabalho, não foram categorizadas como tal por várias razões. Em primeiro lugar, algumas dessas são consideradas parte integrante das estratégias de preservação propriamente ditas (emulação e migração), como é o caso da encapsulação; em segundo lugar, trata-se de abordagens muito complexas e especializadas, que só devem ser utilizadas quando tudo o resto falhou (arqueologia digital), ou, por outro lado, trata-se de abordagens simplistas não aplicáveis a todos os objectos digitais (transferência para suportes analógicos); trata-se de abordagens pouco experimentadas ou inviáveis (não fazer nada).

3.4.1. Encapsulação (*Encapsulation*)

A encapsulação é uma técnica que consiste na criação de uma *cápsula* (ou um pacote de informação, segundo o modelo de referência OAIS), onde se agrupam não só o objecto digital, mas também toda a informação necessária para permitir, no futuro, o desenvolvimento de conversores, visualizadores ou emuladores (Ferreira, 2006, p. 43). Esta é, pois, uma actuação transversal a outras estratégias de preservação anteriormente descritas (como a emulação e a migração *a-pedido*). Por essa razão, é aqui considerada como uma acção inerente a outras estratégias e não como uma estratégia de preservação independente.

Assim encapsulado, o objecto digital e toda a informação a seu respeito ficam em *stand by* até que alguém revele interesse no seu conteúdo. Nessa altura, serão accionados os mecanismos que permitirão aceder ao conteúdo do objecto digital, seja pela criação de emuladores seja pela utilização de conversores ou visualizadores, pois a informação sobre o objecto (encapsulada juntamente com ele) permitirá desenvolver esses mecanismos.

Coloca-se então a grande questão de saber qual a informação sobre o objecto digital que deve ser encapsulada e onde armazenar essa informação e o próprio objecto digital. Lee et al. (2002, p. 98) sugerem:

Encapsulation can be achieved by using physical or logical structures called containers or wrappers to provide a relationship between all information components such as the digital object and some supporting information including metadata. The reference model for the OAIS also describes the types of supporting information that should be included in an encapsulation.

A informação sobre o objecto digital, ou seja, os metadados (de referência, de representação, de proveniência, de fixidez e de contexto) pode ser incluída directamente no objecto digital arquivado ou associada de forma lógica a ele dentro do próprio sistema. Com esta acção, reduz-se a probabilidade de perda de qualquer componente crítico necessário para efeitos de acesso, decodificação e interpretação da informação contida num determinado objecto digital (Kenney et al., 2003). Segundo Webb (2003, p. 128-129), não é prático nem tão-pouco necessário encapsular os meios de acesso utilizados, isto é, os programas e os equipamentos informáticos. Por isso, a encapsulação deve apenas incluir os metadados que descrevem ou estabelecem ligações com as ferramentas necessárias. Um procedimento alternativo é incluir uma especificação do programa ou equipamento informático, ou a ligação a ele, para que o restabelecimento possa efectuar-se no futuro, se necessário.

A grande vantagem desta acção é que ela assegura a manutenção, como um todo, de todos os suportes necessários para garantir o acesso. No entanto, não resolve necessariamente o problema do acesso e, por essa razão, deve ser utilizada como acção complementar de outras estratégias (por exemplo, da emulação). Além disso, nada garante que elementos essenciais não se percam durante a conversão. A encapsulação pode levar à criação de ficheiros muito extensos, que devem ser escritos numa linguagem compreensível por muitas tecnologias e amplamente normalizada (XML, por exemplo). A grande desvantagem da encapsulação está no facto de o *software* encapsulado continuar a estar sujeito a uma rápida obsolescência (Beagrie & Jones, 2008).

3.4.2. Arqueologia digital (*Digital archaeology*)

Por arqueologia digital entende-se todo o processo de recuperação de informação armazenada em suportes danificados, degradados ou obsoletos e/ou formatos obsoletos. Não se trata propriamente de uma estratégia de preservação digital, mas sim de uma abordagem a considerar quando todos os outros esforços de preservação falharem⁶⁴. Este resgate da informação consiste na recuperação dos dados registados nos suportes em forma de *bits* e na

⁶⁴ Keefer & Gallart (2007, p. 91) afirmam mesmo que a arqueologia digital, em vez de uma estratégia, se assemelha a uma ausência de estratégia.

aplicação de medidas que permitam a sua inteligibilidade (Webb, 2003, p. 144-145). Heminger & Robertson (1998) e Ferreira (2006, 2009)⁶⁵, referem-se a esta abordagem como a Pedra de Roseta Digital (*Digital Rosetta Stone*), porque, à semelhança do que aconteceu com a descodificação da Pedra de Rosetta⁶⁶, este processo de recuperação da informação faz-se através da reunião de amostras de informação digital que sejam representativas do formato que se pretende recuperar. Com essa informação será possível inferir das regras necessárias para traduzir a informação contida no suporte e/ou formato originais num formato contemporâneo (Ferreira, 2006, p. 44). Esta abordagem baseia-se no princípio de que a informação preservada em diferentes suportes e em diferentes formatos pode ser usada para recuperar dados armazenados em dispositivos que, entretanto, se tornaram obsoletos ou se degradaram e, através dessa informação, reconstruir os documentos digitais. O seu objectivo é manter o acesso aos documentos digitais no seu formato nativo, sem o converter para os formatos mais recentes e sem necessidade de emular o ambiente tecnológico original.

Defensores da Pedra de Roseta Digital como meio de preservar a informação digital, Heminger & Robertson (1998)⁶⁷ argumentam que a criação duma Pedra de Roseta Digital que possa ser utilizada para recriar os documentos digitais é uma forma de descrever e preservar o comportamento dos sistemas de informação para a posteridade. Assim, para estes autores, esta abordagem constituiria uma alternativa à emulação, na medida em que, à época, não era certo que a informática conseguisse recriar o comportamento do ambiente tecnológico original para, assim, reproduzir os documentos digitais. Além disso, aqueles autores sugerem esta abordagem

⁶⁵ Webb (2003) designa esta abordagem por *Data Recovery*.

⁶⁶ Durante o século XIV, a inexistência de um sistema alfabético que decifrasse os hieroglíficos em que tinha sido registada a maior parte da informação sobre o Antigo Egipto impossibilitou o conhecimento desta civilização que parecia irremediavelmente condenada a uma espécie de agnosia alfabética. No entanto, no final século XVIII, uma expedição napoleónica permitiu descobrir um artefacto que viria a ser conhecido como a Pedra de Roseta. Neste bloco de granito, encontrava-se escrito em três línguas diferentes (egípcio hieroglífico, egípcio cursivo e grego clássico) um decreto emitido por Ptolomeu V Epifânio, em 196 a.C. A informação aí contida foi descodificada por Jean-François Champollion, ainda que qualquer conhecimento sobre a escrita egípcia hieroglífica e cursiva estivesse perdido desde o século IV a.C. Esta descoberta permitiu decifrar a informação sobre o Antigo Egipto.

⁶⁷ O modelo de Pedra de Roseta Digital proposto por estes autores descreve três processos essenciais para manter o acesso à informação digital a longo prazo nos seus formatos nativos: (i) preservação do conhecimento, (ii) recuperação de dados e (iii) a reconstrução dos documentos. O primeiro processo consiste em compilar e preservar grandes quantidades de informação necessárias para recuperar dados digitais de um suporte ou ambiente tecnológico para, assim, reconstruir a informação aí contida. Para tal, seria necessário criar um arquivo de metaconhecimento (conhecimento ou entendimento dos factos, heurística, regras e contexto em que os dados são usados e manipulados). A recuperação de dados consiste na extracção dos dados digitais de um suporte ou ambiente tecnológico degradado ou obsoleto e na sua posterior migração para um suporte e/ou formato actual. Por último, a reconstrução do documento corresponde ao processo de interpretação dos documentos digitais a partir dos dados originais, utilizando a informação sobre o formato do ficheiro compilada durante o processo de preservação do conhecimento.

também como alternativa à migração, pois este tipo de recuperação da informação não implica a sua conversão para formatos mais actuais. A vantagem adicional é óbvia nos casos de colecções com milhões de documentos digitais, em que os custos de migração (pelo menos de dez em dez anos), seriam incomportáveis.

A grande vantagem da arqueologia digital está na sua capacidade de recuperação de dados que, de outro modo, estariam irremediavelmente perdidos. Já foi provado que é possível recuperar grande parte de informação registada em suportes danificados, mas não necessariamente no mesmo formato (Beagrie & Jones, 2008). Ainda assim, não há garantia de que seja possível recuperar a totalidade dos dados, nem a sua legibilidade. Mas existe ainda outra desvantagem: não existindo documentação suficiente, a interpretação dos dados só poderá ser feita por “adivinhação”, tornando-se difícil estabelecer a identidade, integridade e contexto do documento digital recuperado.

Apesar de haver cada vez mais especialistas (“arqueólogos digitais”) a oferecerem este tipo de serviços, a arqueologia digital requer mão-de-obra altamente qualificada e, normalmente, é levada a cabo por organizações privadas de recuperação de dados, que possuem grande variedade de tecnologias e instalações especiais (por exemplo, divisões esterilizadas, onde se desmonta o *hardware* e suportes danificados). Trata-se, naturalmente, de serviços pagos, através dos quais é possível recuperar a cadeia de *bits* dos suportes danificados, sobretudo dos suportes magnéticos. (Kenney et al., 2003). Por ser demasiado dispendiosa, a arqueologia digital é recomendada apenas para a recuperação e restauro de dados contidos em suportes danificados ou formatos obsoletos cujo valor ou importância dos dados justifique esse investimento, pois, por vezes, o valor da informação não justifica o custo da operação (Webb, 2003, p. 145).

3.4.4. Transferência para suportes analógicos⁶⁸

Esta abordagem traduz-se na conversão de documentos digitais em documentos analógicos, como o papel ou o microfilme – que, aliás, têm demonstrado excepcional longevidade –, de forma a garantir o acesso à informação neles contida. Feita a transferência da informação, os esforços de preservação serão feitos em torno do documento analógico. O contra-senso desta abordagem é evidente: por um lado, muitas bibliotecas estão a investir na

⁶⁸ Ferreira (2006) designa esta estratégia por “Migração para suportes analógicos”; Keefer & Gallart (2007) referem-na como “Impressão em suportes analógicos”. Optou-se, neste estudo, por transferência para suportes analógicos com o intuito de, por um lado, não associar esta acção à migração propriamente dita enquanto estratégia de preservação digital; e, por outro lado, não limitar os suportes analógicos apenas ao papel (a transferência da informação digital também pode ser feita para o microfilme, por exemplo).

digitalização como forma de as preservar e tornar acessíveis as suas colecções, simultaneamente, confrontam-se com a necessidade, inversa, de conversão dos documentos nado-digitais em documentos impressos ou microfílmes para assegurar a sua perenidade no tempo. Ainda que Børghoff et al. (2005, p. 39) considerem esta abordagem “deprimente”, a transferência para suportes analógicos pode revelar-se uma solução pragmática e inclusive um último recurso, enquanto se espera pelo desenvolvimento e/ou aplicação de outras estratégias de preservação digital.

No entanto, a transferência da informação de suportes menos estáveis para outros mais estáveis é sustentada pelo facto de o papel *acid-free* de alta qualidade poder durar cerca de 500 anos, enquanto um microfilme pode persistir cerca de 100 anos ou mais (Lazinger, 2001). O papel e o microfilme têm como vantagem adicional o facto de não requererem *hardware* ou *software* específicos para o acesso à informação. Além disso, trata-se de um processo simples e barato: o custo da transferência consiste apenas num acto isolado e único, a conservação da cópia analógica é pouco dispendiosa e as bibliotecas possuem os conhecimentos e procedimentos necessários à sua preservação (Beagrie & Jones, 2008; Keefer & Gallart, 2007, p. 90).

A transferência da informação digital para suportes analógicos é uma estratégia inviável quando se lida com a preservação de objectos digitais dinâmicos e multimédia⁶⁹, pois os suportes analógicos são incapazes de preservar certas funcionalidades dos documentos digitais, como o movimento, o som, as fórmulas, as estruturas de informação, etc. (Saramago, 2002, p. 61). Tendo em conta a vulnerabilidade dos documentos digitais, alguns especialistas defendem a combinação da digitalização com a microfilmagem de documentos analógicos: a cópia digital melhoraria o acesso e a funcionalidade, enquanto a cópia em microfilme funcionaria como um substituto para arquivo (Thomaz e Soares, 2004).

3.4.5. Não fazer nada

Esta hipótese, avançada por Keefer & Gallart (2007, p. 92), assenta na constatação da inércia de muitas bibliotecas e de outras instituições de memória perante a ausência de estratégias e soluções definitivas para preservação da informação digital. No entanto, trata-se de uma hipótese ilógica: tendo em conta os riscos associados aos documentos digitais, a inacção

⁶⁹ Texto e imagens monocromáticas estáticas são os dados mais indicados para este tipo de transferência. Tendo em conta os custos e as limitações dos *backups* analógicos, esta técnica apenas faz sentido para documentos cujo conteúdo mereça um grande nível de redundância e protecção contra a perda. (Kenney et al., 2003)

traduzir-se-ia ainda em maior vulnerabilidade à obsolescência e à degradação. Como foi analisado anteriormente, quanto mais precoce for a intervenção, no caso da preservação digital, maiores são as hipóteses de manter os documentos digitais acessíveis no futuro.

Não fazer nada não deve, portanto, ser uma hipótese a considerar. Perante a incerteza ou impossibilidade de aplicar estratégias de preservação (migração, emulação, encapsulação ou arqueologia digital), é preferível levar a cabo acções básicas: o refrescamento, a preservação em linha ou, em último caso, se possível, a transferência para suportes analógicos.

3.5. Escolha das estratégias de preservação digital

Até à data, não existe uma solução prática aplicável universalmente ao problema da obsolescência tecnológica e da degradação dos objectos digitais. Têm sido propostas várias metodologias, mas é pouco provável que se encontre uma solução única. Nenhuma das estratégias e acções de preservação apresentadas se afigura, por si só, suficiente e completa; possivelmente, será pela combinação de todas que a preservação da informação digital pode ser levada a cabo com sucesso.

Um inquérito levado a cabo por Muir (2004, p. 81) em 2003 revela que as acções e estratégias de preservação digital mais utilizadas pelas bibliotecas são o refrescamento, a preservação da tecnologia, a migração e a emulação. No entanto, segundo o inquérito, muitas bibliotecas optam por transferir para suportes analógicos os documentos digitais, porque, segundo elas, a sua prioridade não é a preservação das cópias digitais.

O importante mesmo, segundo Webb (2003, p. 120), é tomar medidas, ainda que pouco significativas, que preservem o acesso aos objectos digitais. Ao investigar as formas de superar as mudanças tecnológicas, este autor conclui que a maioria das estratégias propostas baseiam-se em pelo menos um dos seguintes princípios: (i) converter os dados numa forma legível para o ser humano, num suporte fácil de manter (por exemplo, papel, microfilme, ...); (ii) criar os dados directamente, ou convertê-los depois, para forma altamente normalizada de codificação ou formato, ou ambas, que possa ser amplamente reconhecida pelos sistemas informáticos durante muito tempo; (iii) fazer com que os dados sejam “autodescritivos” e “autosustentáveis”, incorporando-lhes metadados e ligações com os programas informáticos que os mantêm acessíveis durante algum tempo (e talvez incorporando-lhes também outros programas); (iv) converter os dados para um formato que permita encontrar mais facilmente os meios de acesso; (v) manter os dados na sua forma original e fornecer as ferramentas que os representam nessa forma, seja através dos programas e equipamentos originais (também eles preservados), seja através de novos programas que emulem o comportamento destes; (vi) prever especificações

para emular os meios de acesso originais numa plataforma informática intermédia teórica, que funcionará como ponte para uma emulação posterior num novo ambiente tecnológico; (vii) migrar os dados para os novos formatos acessíveis com cada nova tecnologia; (viii) permitir uma migração posterior *a-pedido*, mantendo os dados e registando informação suficiente sobre estes para que um futuro utilizador ou responsável possam convertê-los de forma compreensível no momento desejado; (ix) manter os dados e fornecer novos programas informáticos de apresentação que produziriam uma apresentação aceitável dos dados em cada novo ambiente tecnológico.

A escolha das acções e/ou estratégias de preservação que garantam o acesso aos objectos digitais será sempre condicionada pela natureza da organização que as executa, pelas características dos objectos digitais, pelos custos associados e, sobretudo, pelas necessidades e exigências da comunidade de interesse.

EMULAÇÃO VS MIGRAÇÃO

Neste trabalho, apenas se consideraram a emulação e a migração como estratégias efectivas de preservação da informação digital não só pela porque ambas assumem proeminência neste contexto mas também porque se revelam as acções mais capazes. O valor relativo destas duas estratégias parece ter dominado a discussão, desde meados dos anos 90, sobre que estratégia de preservação utilizar. A mais-valia de cada uma delas foi medida em função da sua capacidade de manter a aparência e funcionalidade ou da sua capacidade de preservar apenas as propriedades significativas. Os defensores da emulação argumentam que esta estratégia é a única que garante a preservação do documento digital, mantendo-se a sua aparência e funcionalidade, pois converte o ambiente tecnológico que permite lê-lo no seu estado original. Os entusiastas da migração invocam o facto de esta estratégia permitir converter o documento original para que este funcione nos sistemas novos (mantendo as suas propriedades significativas), o que torna o processo de recuperação e acesso ao item mais rápido.

No entanto, ambas as estratégias se encontram sob a mesma ameaça: a obsolescência. À semelhança do que acontece com os emuladores, também os formatos de destino do processo de migração tornar-se-ão, mais cedo ou mais tarde, obsoletos. Deste modo, a emulação da tecnologia, bem como a migração da informação, pode tornar-se num “*cuanto de nunca acabar*”, de emular a emulação da emulação, de migrar a migração da migração... (Bia e Sánchez, 2002).

A adopção de uma ou outra estratégia (ou de ambas) deve resultar de uma decisão ponderada e baseada em vários factores, nomeadamente o tipo de objecto digital a preservar e quais as propriedades do objecto que se pretende preservar:

The issues of migration and emulation no longer dominate the agendas of meetings and conferences. Most people involved in making decisions about digital collections are comfortable with the notion that it will be necessary to take one approach or the other, and they are content to make that decision when the time comes. They also recognize that that decision will be made more than once, and can be remade as required, at least for the first generation of changes. Major or quantum changes, of course, may well demand a definitive and final decision, but the current range of incremental technological changes means that the time of the disruptive and irreversible change is not yet here. (Bradley, 2007, p. 154)

O foco da preservação deslocou-se, neste debate, da preocupação em preservar a cadeia de *bits* que constituem o objecto digital para a preocupação em preservar o seu conteúdo. A comunidade mais pragmática da preservação digital concentrou-se nesta nova perspectiva. É preciso, antes de mais, preservar o acesso e não apenas o objecto digital ou suporte que o contém. O acesso corresponde não só à capacidade de encontrar e recuperar um item, mas também à capacidade de usar, ver, ouvir, interagir, exibir e fazer correr o objecto digital para que os utilizadores possam satisfazer as suas necessidades informativas (Bradley, 2007, p. 154).

INDICAÇÕES DE USO

Diferentes tipos de acções/estratégias são viáveis para objectos digitais de diferentes tipos. Não existe nenhuma receita para a escolha da(s) melhor(es) estratégia(s) a aplicar. Esta terá de ser sempre uma escolha ponderada, que analise o tipo de objecto digital⁷⁰ a preservar e as vantagens e desvantagens de cada acção/estratégia. Webb (2003) e Lee et al. (2002) sugerem o domínio de uso de cada uma delas:

⁷⁰ Lee et al. (2002, p. 102), apresentam um diagrama de apoio à selecção das estratégias de preservação que pode ser consultado no Anexo 4.

Tabela 1: *Ações e estratégias de preservação digital e suas indicações de uso.*

Refrescamento	<ul style="list-style-type: none"> – Acção transversal a todas as outras.
Preservação da tecnologia	<ul style="list-style-type: none"> – Medida a curto prazo que deve ser levada a cabo até que outras acções sejam executadas; – Destina-se, sobretudo, a objectos digitais complexos, como programas e objectos multimédia; – Recomenda-se também para os programas informáticos necessários para que algumas estratégias possam ser implementadas.
Emulação	<ul style="list-style-type: none"> – Objectos digitais muito complexos ou programas executáveis que só podem funcionar em ambientes tecnológicos específicos; – Objectos digitais cujo valor reside na sua visualização no ambiente tecnológico original; – Se já existem emuladores adequados para as plataformas necessárias ou caso se disponha dos conhecimentos necessários para construir o emulador.
Migração	<ul style="list-style-type: none"> – Objectos digitais relativamente simples e que não contenham ficheiros executáveis e cuja visualização não dependa da sua aparência e funcionalidade; – Objectos digitais activamente acedidos e manuseados; – Ideal para colecções digitais homogéneas que se encontrem em formatos normalizados e não sujeitos a propriedade intelectual.
Encapsulação	<ul style="list-style-type: none"> – Objectos digitais relativamente simples e em formatos normalizados; – Objectos cujo acesso e gestão activa são pouco prováveis.
Arqueologia Digital	<ul style="list-style-type: none"> – Recomenda-se unicamente para a recuperação e restauro de dados em caso de suportes danificados ou formatos obsoletos ou caso o valor ou importância dos dados justifique o custo.
Transferência para suportes analógicos	<ul style="list-style-type: none"> – Objectos digitais que não requeiram a conservação das funcionalidades tecnológicas (elementos multimédia, hiperligações, dados dinâmicos...) para atingir os seus fins, ou seja, objectos digitais estáticos (documentos de texto, imagem, etc).

A escolha de acção/estratégia a aplicar deve ter em conta a dimensão do objecto digital que se pretende preservar. As opções variam entre: (i) preservar a sua dimensão física (acondicionamento, manuseamento e boas práticas adequados, refrescamento e preservação da tecnologia) para garantir a sua preservação das dimensões lógica e conceptual; (ii) preservar a sua dimensão lógica quando as acções anteriores não se revelem eficazes (emulação, encapsulamento) e (iii) preservar a sua dimensão conceptual (migração), ou seja, apenas o conteúdo objectivo último de todas as acções/estratégias.

4. Repositórios digitais e a preservação digital

Como foi referido no capítulo um deste trabalho, os repositórios digitais⁷¹ apresentam-se actualmente como uma ferramenta crítica e central no processo de preservação da informação digital. Márdero Arellano (2008, p. 124) distingue repositórios temáticos (cobrem uma determinada área do conhecimento), repositórios institucionais (tornam acessível a produção intelectual de instituições e comunidades científicas) e repositórios centrais (normalmente são provedores de serviços nacionais e internacionais responsáveis pela colecta de documentos que se encontram em bibliotecas digitais, bases de dados e repositórios temáticos ou institucionais). Tanto os repositórios temáticos como os repositórios institucionais surgem como meios de divulgação e comunicação da produção científica, paralelos, por exemplo, às revistas científicas. A sua missão é, pois, divulgar e tornar acessível um tipo de informação específico – a informação científica. Naturalmente, os repositórios temáticos e os repositórios institucionais devem assumir o compromisso de preservação do material digital que colectam⁷², nem que seja apenas a curto/médio prazo. Apenas os repositórios centrais (com um âmbito mais alargado dado o seu cariz nacional ou internacional) devem assumir missão a preservação digital a longo prazo. Idealmente, este tipo de repositórios digitais permite armazenar os documentos digitais, mantê-los acessíveis, migrá-los quando necessário, documentar as alterações feitas ao longo do processo de custódia e preservação e manter a autenticidade dos documentos custodiados.

⁷¹ Na literatura, muitas vezes utiliza-se indistintamente os termos “repositório digital” e “arquivo digital” (o modelo de referência OAIS, por exemplo, utiliza o termo “arquivo” para referir a organização que preserva a informação de forma a mantê-la acessível e utilizável por uma Comunidade-Alvo). Neste trabalho, adoptou-se o termo “repositório digital” em detrimento de “arquivo digital”, porque, tal como Barbedo et al. (2007), entende-se que um arquivo digital corresponde a uma estrutura que compreende tecnologia, recursos humanos e um conjunto de políticas para incorporar, gerir e tornar acessíveis, numa perspectiva continuada, objectos digitais de natureza arquivística (a informação de arquivo distingue-se de qualquer outra pelo facto de ser produzida com o propósito primário de constituir prova de uma actividade organizacional). Por essa razão, não se deve confundir repositórios digitais com arquivos digitais; da mesma maneira, segundo Kenney et al. (2008) e Márdero Arellano (2008, p. 124), não se deve confundir repositórios digitais com bibliotecas digitais. Se, por um lado, é a natureza da informação dos arquivos digitais que os distingue dos repositórios, o que distingue as bibliotecas digitais dos repositórios é que as primeiras apenas colecionam e providenciam o acesso à informação digital sem que se comprometam a preservá-la e mantê-la acessível a longo prazo. Assim, os repositórios digitais realizam algumas funções das bibliotecas digitais (seleccionam, obtêm, armazenam e tornam acessível a informação digital). Todavia, no que concerne à preservação dessa informação, os repositórios são considerados parte da infra-estrutura dos serviços de uma biblioteca digital. A natureza da informação, por um lado, e a função de preservação, por outro lado, parecem ser aquilo que, respectivamente, distingue os repositórios digitais de arquivos digitais e de bibliotecas digitais.

⁷² Esta questão não é consensual na literatura.

(...) digital repositories will function not just as guarantors of the long-term viability of materials in their custody, but also as access gateways. Fulfilling this dual mission requires that preservation processes operate seamlessly alongside access services. Preservation should not impede access or reduce the scope for sharing information. Digital content management systems must find ways to integrate preservation tools and services into their environments. (Lavoie ; Dempsey, 2004)

Segundo Owen (2007, p. 49), perante a dificuldade e/ou incapacidade das instituições de memória tradicionais (bibliotecas, arquivos, centros de documentação, museus...), os repositórios digitais surgem como uma nova infra-estrutura que visa preservar e tornar acessível o património digital. Não descurando a importância daquelas instituições na preservação da memória colectiva pré-digital, a nova entidade deve ser complementar à actividade dessas vetustas instituições de memória e, assim, responsabilizar-se pela preservação do património cultural em formato digital. Os repositórios digitais não substituem as bibliotecas ou os arquivos, eles estendem significativamente os seus papéis e representam um sério compromisso com a preservação da informação em formato digital. Deste modo, os repositórios digitais podem ser considerados as instituições de memória da sociedade digital. Webb (2003, p. 38-39) apelida-os de *safe places* onde o património digital deve ser alojado e onde pode ser protegido de todas as ameaças a que os seus componentes físicos e lógicos estão expostos, permanecendo acessível a longo prazo.

Um repositório digital é um sistema de informação responsável por gerir e armazenar informação digital (Ferreira, 2009, p. xxvii), mantendo-a acessível e preservando-a. Segundo Wheatley (2004, apud Hockx-Yu, 2006, p. 4), preservar, num repositório digital, significa que se cumprem os seguintes objectivos: (i) os dados são mantidos no repositório sem serem danificados, perdidos ou adulterados; (ii) os dados podem ser encontrados, recuperados e disponibilizados ao utilizador; (iii) os dados podem ser interpretados e entendidos pelo utilizador (não basta preservar a *bitstream* que representa a informação preservada, o desafio reside em assegurar que os utilizadores podem aceder ao conteúdo tal como este foi depositado no repositório; (iv) todas as condições referidas podem ser mantidas ao longo do tempo.

Assim, no contexto da preservação digital, um repositório digital surge como uma importante infra-estrutura que providencia os mecanismos necessários para (i) identificar de forma consistente os documentos, (ii) uniformizar os formatos em que estes se encontram e (iii) atribuir-lhes os metadados necessários para a sua descrição, localização, gestão dos direitos de autor e documentação das alterações efectuadas aos documentos. Em termos de preservação, essas acções garantem uma melhor gestão das mudanças tecnológicas e da aplicação das estratégias de preservação, com a automatização de processos. Por exemplo, a migração de conteúdos digitais torna-se mais simples, mais rápida e menos dispendiosa quando aplicada a um conjunto de documentos com o mesmo formato.

REPOSITÓRIOS DIGITAIS DE CONFIANÇA (TRUSTED DIGITAL REPOSITORIES)

Uma noção central no âmbito da preservação “confiança”. Em virtude da instabilidade e vulnerabilidade dos documentos digitais, os seus curadores devem assegurar que os documentos sejam preservados tal qual foram depositados, ou seja, que se mantêm a sua autenticidade e integridade e que todas as operações de preservação levadas a cabo estão devidamente justificadas e documentadas.

A consciência da necessidade de repositórios digitais de confiança surge num documento de 2002 publicado pelo *Research Libraries Group* (RLG) em parceria com a *Online Computer Library Center* (OCLC) intitulado *Trusted Digital Repositories: attributes and responsibilities*. Neste relatório, baseado no modelo de referência OAIS, caracterizam-se os repositórios digitais (identificando os seus atributos e responsabilidades⁷³) cuja missão é a manutenção e gestão de colecções digitais, preservando-as e tornando-as acessíveis de forma autêntica e fiável a uma Comunidade-Alvo de consumidores:

A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. Trusted digital repositories may take different forms: some institutions may choose to build local repositories while others may choose to manage the logical and intellectual aspects of a repository while contracting with a third-party provider for its storage and maintenance. (RLG/OCLC, 2002, p. 5)

Os repositórios digitais são infra-estruturas recentes e a base de confiança que lhes é exigida (a mesma que bibliotecas, arquivos e museus sedimentaram ao longo de décadas), não é um objectivo que se estabeleça numa década. Assim, para que a confiança se estabeleça entre os seus utilizadores, RLG/OCLC recomendam que todos os repositórios digitais ditos de confiança sejam sujeitos a um processo de certificação:

For organizations wishing to provide digital repository services, building the required trust via reliable, proven practice will take time. Yet the need for trusted services is immediate. Since action needs to be taken now to preserve the already large body of digital materials, a program for certification should be developed to provide a basis for trustworthiness. Certification would dictate criteria that must be met and would employ mechanisms for assessment and measurement. (RLG/OCLC, 2002, p. 37)

⁷³ Ver Anexo 5

4.1. Modelo de referência OAIS

Aprovado como norma internacional em 2003⁷⁴ – ISO Standard 14721:2003 –, o Modelo de Referência OAIS⁷⁵ (*Open Archival Information System*) nasce do esforço conjunto do *Consultative Committee for Space Data Systems* (CCSDS) e da *International Organization of Standardization* (ISO), encetado em 1990, com o objectivo de regular o armazenamento da informação digital a longo prazo.

Nesta norma, encontram-se especificadas as recomendações técnicas para o desenvolvimento de um arquivo⁷⁶ que permita preservar a longo prazo e manter acessível a informação digital. Assim, um arquivo, segundo o modelo de referência OAIS, consiste numa organização de pessoas e sistemas que têm como principal responsabilidade preservar a informação digital e torná-la acessível a uma Comunidade-Alvo. Esta definição enfatiza as duas principais funções de um arquivo deste tipo: (i) preservar a informação, assegurando a sua persistência a longo prazo e (ii) tornar a informação arquivada acessível de forma inteligível e de acordo com as necessidades dos consumidores do modelo de arquivo OAIS.

De acordo com o CCSDS (2002, p. 3-1), as responsabilidades mandatárias de um arquivo segundo o modelo de referência OAIS são: (i) negociar e aceitar a informação adequada dos Produtores e/ou detentores dos direitos da informação; (ii) obter o controlo efectivo da informação como garantia da sua preservação a longo prazo; (iii) determinar a Comunidade-Alvo capaz de compreender a informação fornecida; (iv) garantir que a informação está acessível e é compreendida pela Comunidade-Alvo sem que esta tenha de recorrer aos produtores da informação; (v) seguir políticas e procedimentos documentados que garantam que a informação é preservada e pode ser disseminada enquanto cópia autêntica do original; (vi) tornar a informação preservada disponível à Comunidade-Alvo.

Trata-se de um sistema de informação dedicado à preservação digital (Ferreira, 2006, p. 27), onde são apresentadas e definidas as entidades funcionais que o compõem. Lavoie (2004, p.

⁷⁴ Embora se previsse a revisão deste documento cinco anos após a sua publicação, só em 2009 foi publicado um *draft* dessa revisão, aguardando-se, até à data a sua publicação oficial. Para este trabalho, consultou-se os dois documentos, pois o *draft* apresenta novos e importantes desenvolvimentos no que concerne, por exemplo, à autenticidade e aos direitos de acesso.

⁷⁵ Não confundir com o acrónimo OAI (*Open Archives Initiative*) que se refere à iniciativa internacional que visa a interoperabilidade e rápida disseminação da informação, facilitando o acesso aos documentos digitais depositados nos diferentes repositórios.

⁷⁶ O modelo de referência OAIS utiliza a expressão “arquivo digital” para designar a organização responsável pela preservação digital. Neste subcapítulo, utilizaremos o termo “arquivo”, em vez de “repositório” de forma a respeitar a terminologia utilizada na norma.

5) refere que um modelo OAIS não opera no vácuo. Na verdade, existem entidades funcionais externas e internas que o tornam operacional:

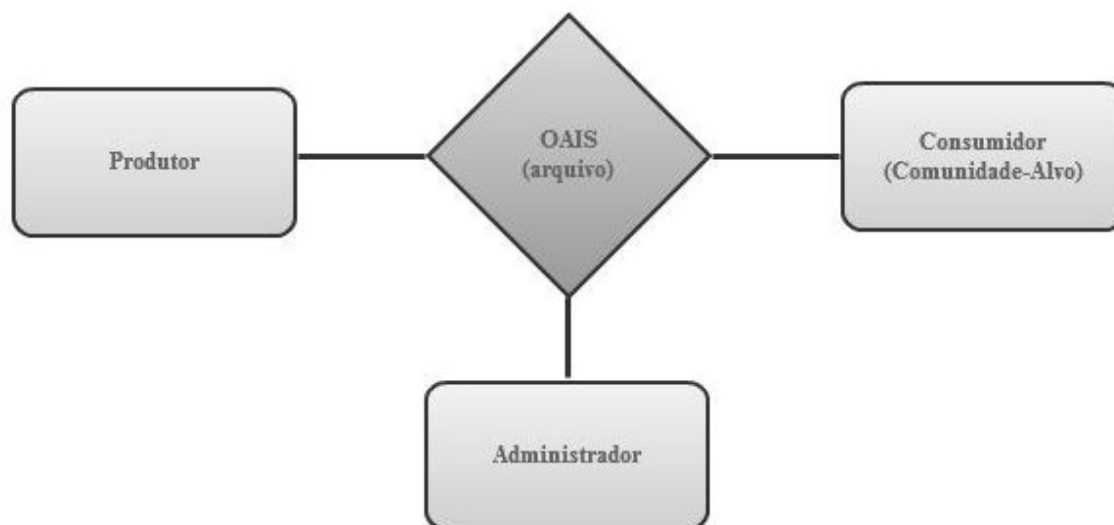


Figura 10 – Entidades externas de um arquivo OAIS.

O Produtor corresponde à(s) pessoa(s) física(s) ou organização(ões) que fornece(m) ao modelo OAIS a informação a ser preservada ao longo do tempo. O Administrador é responsável pela política estrutural de alto nível das actividades do OAIS, tendo a responsabilidade de, entre outros aspectos, (i) definir o planeamento estratégico; (i) rever periodicamente as políticas do OAIS; (iii) avaliar o desempenho do OAIS. O Administrador não se envolve em actividades de gestão diárias das operações do OAIS (esta função é desempenhada por uma função dentro do próprio arquivo – Administração). O Consumidor inclui os indivíduos, organizações ou sistemas que podem aceder e utilizar a informação arquivada no OAIS. Dentro deste grupo, o modelo de referência OAIS distingue a Comunidade-Alvo como o grupo especial de consumidores aptos a compreenderem a informação preservada de forma autónoma e independente.

A submissão da informação por parte do Produtor, o armazenamento, gestão e preservação dentro do Arquivo e a disseminação da informação por parte do Consumidor são feitos através de Pacotes de Informação (PI), os quais, dependendo do momento e do lugar em que se encontram dentro do arquivo, assumem diferentes tipologias: Pacote de Informação para Submissão (PIS), Pacote de Informação para Arquivo (PIA), Pacote de Informação para Disseminação (PID) ⁷⁷.

⁷⁷ Ver Anexo 6

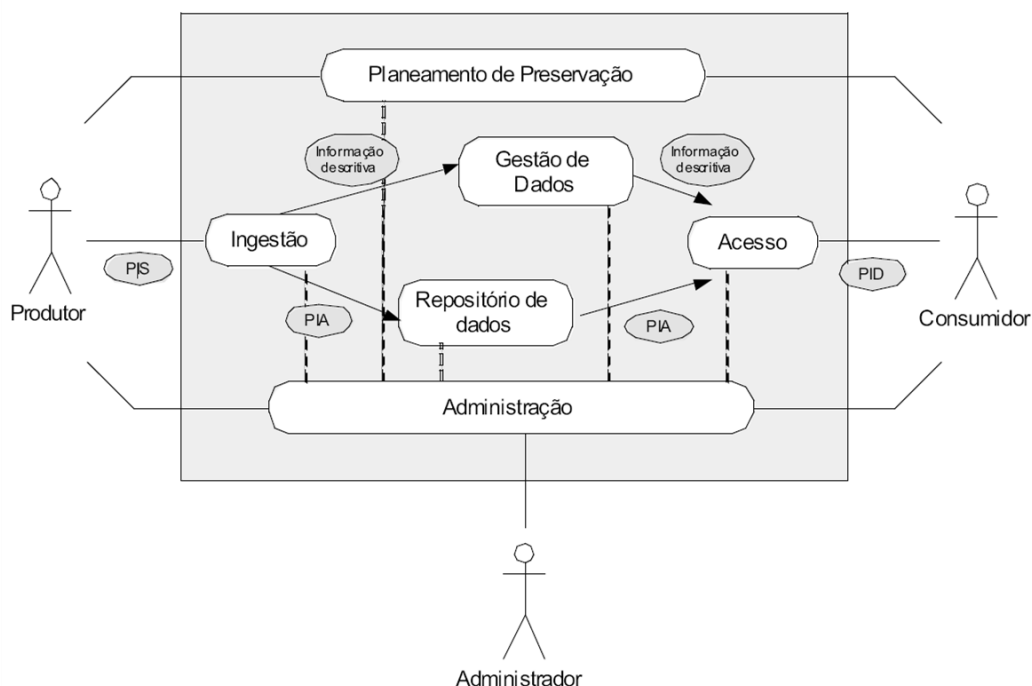


Figura 11- Modelo funcional do OAIS. Fonte: Ferreira (2006, p. 28)

São seis as entidades internas que tornam o modelo operacional: a Ingestão, o Repositório de Dados, a Gestão de Dados, a Administração, o Planeamento da Preservação e o Acesso. Para uma descrição sucinta da forma como a informação pode ser preservada num arquivo digital compatível com o modelo de referência OAIS, é necessário verificar a forma como as entidades externas interagem com as entidades funcionais internas e a forma como o PI vai sendo tratado e se vai transformando durante o processo de ingestão, arquivo e acesso.

Em primeiro lugar, ocorre a actuação do Produtor, entidade externa ao repositório, que submete um objecto digital e os metadados associados para arquivo, usando, para tal, um PIS. A submissão da informação é feita através da **Ingestão**, entidade interna que serve de interface entre o Produtor e o OAIS. Num arquivo segundo o modelo OAIS, a Ingestão é responsável por, entre outros aspectos: (i) aceitar o PIS enviado pelo Produtor e gerir todo o processo de aceitação da custódia dessa informação; (ii) transformar o PIS em PIA e enviá-lo para o Repositório de Dados; (iii) extrair a informação descritiva que servirá de suporte à pesquisa e enviá-la para o Gestão de Dados. Toda a submissão de informação deve ser formalizada através de Acordo de Submissão, que estabelece detalhes específicos, como o tipo de informação submetida, os metadados fornecidos pelo Produtor e a transferência da custódia do Produtor para o Arquivo.

No momento da ingestão, verificam-se os dados, geram-se PIAs (de acordo com a formatação de dados e os padrões de documentação do arquivo) e produz-se toda a Informação Descritiva (ou seja, os metadados) que, posteriormente, suportará o acesso ao objecto digital

depositado. Os PIAs são enviados para o **Repositório de Dados**, o qual, dentro de um arquivo OAIS, é a entidade responsável por fornecer os serviços e funções para o armazenamento, manutenção e recuperação da informação, tais como: (i) receber o PIA e depositá-lo no arquivo permanente; (ii) refrescar e/ou migrar a informação para outros suportes e/ou formatos; (iii) planejar formas de recuperação de catástrofes de forma a mitigar os seus efeitos; (iv) recuperar os PIAs de modo a satisfazer os pedidos de acesso por parte do consumidor. Por seu turno, a Informação Descritiva gerada, juntamente com outros metadados necessários à preservação e ao acesso à informação, é enviada para a entidade **Gestão de Dados**, que tem como responsabilidades: (i) criar e manter bases de dados de metadados descritivos para a identificação e descrição da informação armazenada no arquivo permanente de modo a suportar as ajudas de pesquisa do OAIS; (ii) fazer pesquisas nestas bases de dados, avaliar os resultados e gerar relatórios; (iii) actualizar os dados sempre que sejam depositados novos PISs ou sempre que os PIAs existentes sejam modificados ou apagados.

À semelhança da Ingestão, o **Acesso** é a entidade que serve de interface entre o OAIS e os seus Consumidores, nomeadamente a Comunidade-Alvo, uma vez que é a entidade responsável por gerir os processos e serviços através dos quais os Consumidores (em particular, a Comunidade-Alvo), localizam, solicitam e acedem à informação custodiada pelo OAIS. Cabe ao Acesso: (i) processar os pedidos feitos ao OAIS, devolvê-los à Gestão de Dados e apresentar os resultados ao Consumidor, sob a forma de PID; (ii) implementar mecanismos de controlo da segurança e acesso aos conteúdos sob a custódia do OAIS. O Acesso é o mecanismo através do qual o OAIS cumpre a responsabilidade de tornar a informação nele depositada e por ele custodiada disponível e acessível aos consumidores. Através da Informação Descritiva e dos PIAs, o Acesso facilita a descoberta e a localização dos objectos digitais pretendidos pelo Consumidor.

O **Planeamento de Preservação** é responsável pela definição das políticas de preservação que permitem manter a informação acessível, recorrendo, para tal, a estratégias de preservação da informação digital. É a entidade responsável por, entre outros aspectos: (i) analisar, no ambiente externo, mudanças que possam causar impacto na capacidade de o OAIS em preservar e manter o acesso à informação sob sua custódia; (ii) desenvolver e aplicar estratégias de preservação da informação; (iii) estar atento ao desenvolvimento constante da tecnologia, de forma a poder actualizar políticas de preservação e protecção da informação, (iv) elaborar relatórios periódicos com a análise de riscos.

Por último, embora transversal a todas as outras, a **Administração** é a entidade interna responsável por gerir as actividades quotidianas do OAIS e por coordenar as actividades das outras cinco entidades internas. É da responsabilidade da Administração: (i) interagir com as três entidades externas, isto é, com os Produtores (através da formulação dos Acordos de

Submissão, por exemplo), com os Consumidores (através de serviços de apoio ao consumidor, por exemplo) e com o Administrador (com a definição, implementação e manutenção de políticas); (ii) supervisionar as operações dos sistemas de arquivo e acesso; (iii) verificar o desempenho do sistema e coordenar as actualizações sempre que necessário.

Em suma, o modelo OAIS abrange três entidades externas (Produtor, Consumidor e Administração) e seis entidades funcionais internas (Ingestão, Repositório de Dados, Gestão de Dados, Acesso, Planeamento de Preservação e Administração) que, em conjunto e em interacção, constituem o mecanismo através do qual um arquivo que segue o modelo OAIS preserva a informação a longo prazo e a torna acessível a uma dada comunidade (Lavoie, 2004., p. 10).

4.2. Metadados de preservação

O modelo de referência OAIS, ao detalhar sobre os tipos de informação necessária para a preservação dos objectos digitais (Informação Descritiva de Preservação⁷⁸), constitui, de certo modo, o ponto de partida para a reflexão sobre que metadados são necessários para assegurar a preservação e acesso dos objectos digitais a longo prazo.

Tradicionalmente, os metadados são genericamente descritos como sendo dados sobre os dados. Esta definição radica na própria etimologia da palavra, em que o prefixo *meta-* indica que algo se aplica a si mesmo. Esta definição simplista de metadados, está na base de definições mais completas:

Metadados são elementos ou estruturas de organização da informação que, associados a um objecto digital, classificam, categorizam e descrevem essa informação. Em suma, são dados sobre os dados, são a informação estruturada sobre a informação distribuída. (Méndez Rodríguez, 2002, p. 29)

Metadado é toda aquela informação que descreve o contexto, qualidade, condição ou características de um recurso, dado ou objecto que tem a finalidade de facilitar a sua recuperação, autenticação, avaliação, preservação ou interoperabilidade. (Senso, 2003, p. 99)

Metadados correspondem à informação estruturada que descreve, explica, localiza, ou de qualquer outro modo, facilita a recuperação, uso ou gestão de um de um recurso de informação. Geralmente, são designados como dados sobre os dados ou informação sobre a informação. A catalogação tradicional que se faz nas bibliotecas é uma forma de metadados; os registos MARC 21 e outros conjuntos de regras utilizados como o AACR2, são esquemas de metadados. (NISO, 2004)⁷⁹

⁷⁸ Ver Anexo 6.

⁷⁹ Esta definição remete para o facto de os metadados não serem exclusivos dos documentos digitais, embora algumas definições se referirem a eles como um fenómeno recente e exclusivo deste tipo de

Os metadados podem estar embutidos dentro do próprio objecto digital (através de um cabeçalho HTML, por exemplo) ou podem corresponder a um registo separado (um registo em formato MARC, por exemplo). O armazenamento dos metadados juntamente com o objecto que eles descrevem garante que os metadados não se perderão; resolve os problemas de ligação entre dados e metadados; contribui para assegurar que tanto os dados como os metadados são actualizados em conjunto. Por outro lado, armazenar os metadados separados dos dados pode simplificar a gestão dos próprios metadados e facilitar a pesquisa e a recuperação da informação.

Os metadados são de três tipos⁸⁰: metadados administrativos, metadados descritivos, metadados estruturais.

Os metadados administrativos são criados com objectivos de gestão, tomada de decisão e armazenamento do objecto digital. Fornecem informação sobre os requisitos de armazenamento e sobre os processos de migração dos objectos digitais. Servem para apoiar as actividades de monitorização, reprodução e *backup* dos pacotes de informação. Incluem informação sobre a aquisição/proveniência (como e quando foi criado o pacote de informação, quando foi adquirido, modificado, etc.), sobre os direitos de autor, sobre permissões de uso e de reprodução e sobre os requisitos legais de acesso, sobre a localização e identificação do objecto digital e sobre a preservação⁸¹.

documentos Na verdade, a sua utilização já se verificava, mas não com esta designação, na descrição bibliográfica tradicional. O argumento de que os metadados são exclusivos dos objectos digitais na medida em que se encontram encerrados dentro do próprio objecto também é facilmente refutável: no mundo analógico, os metadados também podem estar integrados no próprio recurso. A CIP (*Cataloguing in Publication*) é disso exemplo. Por sua vez, os metadados separados dos recursos, se eles existem hoje no mundo virtual através do OPACs e das bases de dados, eles existiram primeiramente nos catálogos tradicionais ou nos índices, muito antes de estes serem automatizados.

⁸⁰ Dada a inexistência de uma taxonomia formal, a categorização dos metadados varia de autor para autor. Assim, aquilo que uns autores designam por metadados estruturais, outros podem designar por metadados de utilização; e aquilo que uns autores constitui uma categoria à parte (metadados de preservação) pode ser para outros como uma subcategoria (como parte integrante dos metadados administrativos, por exemplo).

⁸¹ Taylor (2004) subdivide os metadados administrativos em: metadados de preservação, metadados de acesso/direitos, metametadados. Os metadados de preservação correspondem à informação necessária para assegurar o armazenamento e usabilidade a longo prazo do conteúdo digital (informação sobre os processos de preservação utilizados – reformatação, migração, emulação, integridade do ficheiro, proveniência,... -, descrição do ficheiro, tamanho, propriedades, ambiente tecnológico, fonte de informação, historial do objecto,...). Os metadados de acesso/direitos correspondem à informação sobre quem tem acesso aos pacotes de informação, quem pode utilizá-los, em que circunstâncias e para que fins. Lidam com a propriedade intelectual e *copyright* e com os acordos legais que permitem aos utilizadores aceder à informação e à instituição que os tem sob custódia preservá-los. Os metametadados são, como a própria palavra indica, os dados sobre os metadados, ou seja, correspondem à informação sobre os próprios metadados. A sua principal função é assegurar a autenticidade dos metadados.

Os metadados estruturais correspondem à composição ou estrutura do ficheiro, ou seja, correspondem ao pacote de informação que está a ser descrito, isto é, à informação técnica necessária para assegurar que um pacote de informação funciona correctamente. Os metadados estruturais informam sobre o modo como os ficheiros relacionados permanecem juntos e como um objecto pode ser exibido e disseminado nos vários sistemas. Lidam com aquilo que um objecto é, com aquilo que ele faz e como funciona (por isso, são também designados por metadados técnicos, metadados de exibição ou metadados de uso). Incluem informação sobre o *software* e o *hardware*, informação técnica (tamanho dos ficheiros, comprimento de *bit*, formato, regras de apresentação, sequência da informação, mapas estruturais, compressão do ficheiro, ...), controlo de versões, informação sobre a digitalização (dados relacionados com a criação da imagem digitalizada – scanner utilizado, por exemplo), informação sobre autenticação e dados de segurança, protocolos de pesquisa associados (Z39.50).

Os metadados descritivos equivalem aos dados que descrevem as características que identificam um pacote de informação em conjunto com a análise do seu conteúdo intelectual. Incluem dados que identificam o pacote de informação (título, autor, data de criação/publicação,...), dados sobre a organização intelectual (controlo de autoridade, assunto(s),...), dados sobre o acesso intelectual (cabecinhos de assunto, classificação, categorização,...). Em suma, metadados administrativos documentam os dados de gestão ao longo do tempo; os metadados estruturais, complementam a informação administrativa pois acrescentam o enquadramento tecnológico indispensável à recuperação dos recursos; e os metadados descritivos estão para os recursos digitais como os formatos MARC estão para os documentos tradicionais (Saramago, 2004).

A par destes três tipos de metadados, alguns autores (Webb, 2003) defendem a existência de uma outra classe de metadados: os metadados de preservação. Este tipo de metadados também é considerado por alguns autores (Taylor, 2004) como uma subclasse dos metadados administrativos, uma vez que a preservação é uma das responsabilidades da gestão da informação. Atente-se as seguintes definições de metadados de preservação:

Metadados de preservação correspondem à informação estruturada sobre o objecto digital que: (i) identifica o material sobre o qual o programa de preservação tem responsabilidade; (ii) indica o que é necessário para manter e proteger os dados; (iii) indica ao utilizador aquilo que é necessário para representar o objecto digital pretendido, independentemente das evoluções que tenham ocorrido nas tecnologias de armazenamento e acesso; (iv) regista o histórico do objecto, bem como as consequências das operações que tem sofrido; (v) documenta a identidade e a integridade do objecto para certificar a sua autenticidade; (vi) permite ao utilizador e ao programa de preservação compreenderem o contexto do objecto digital armazenado.
(Webb, 2003, p. 100)

Preservation metadata is information that supports and documents the long-term preservation of digital materials. It addresses an archived digital object's provenance, documenting the custodial history of the object; authenticity, validating that the digital object is in fact what it purports to be, and has not been altered in an undocumented way; preservation activity, documenting the actions taken to preserve the digital object, and any consequences of these actions that impact its look, feel, or functionality; technical environment, describing the technical requirements, such as hardware and software, needed to render and use the digital object; and rights management, recording any binding intellectual property rights that may limit the repository's ability to preserve and disseminate the digital object over time. Preservation metadata addresses all of these issues and more. In short, preservation metadata helps make an archived digital object self-documenting over time, even as the intellectual, economic, legal, and technical environments surrounding the object are in a constant state of change. (Lavoie & Gartner, 2005, p. 1)

O tipo de informação que, segundo estas definições, os metadados de preservação devem incorporar corresponde a alguma da informação que os outros três tipos de metadados já integram. Assim, actualmente tende reconhecer-se os metadados de preservação como uma categoria de metadados à parte, que engloba as demais (Keefer & Gallart, 2007, p. 140): como explicam Lavoie & Gartner (2005, p. 5), os metadados de preservação são metadados descritivos, estruturais e administrativos que apoiam a preservação a longo prazo dos materiais digitais.



Figura 12: Metadados de preservação.

Várias instituições e iniciativas internacionais⁸² têm constituído importantes marcos na identificação dos elementos específicos que devem incluir um esquema⁸³ de metadados de

⁸² Modelo de referência OAIS, PREMIS, CEDARS, NEDLIB...

preservação que reflecta as inúmeras necessidades e funções a ser resolver no âmbito da preservação da informação digital. No entanto, não é objectivo deste trabalho descrever pormenorizadamente essas iniciativas e os esquemas que delas surgiram. Interessa antes salientar a importância e principais funções dos metadados de preservação.

Caplan (2009) apresenta alguns exemplos ilustrativos das actividades de preservação e a forma como podem ser suportadas pelos metadados: (i) um recurso deve armazenar-se de forma segura para que nada possa modificar inadvertida ou intencionalmente o recurso (a informação arquivada como metadados pode ser utilizada para verificar se o item armazenado foi alterado em algum momento), (ii) os ficheiros devem ser armazenados em suportes que sejam legíveis pelos ambientes tecnológicos actuais, de forma a evitar a sua degradação ou obsolescência (os metadados podem apoiar a gestão dos diferentes suportes através do registo do tipo e idade do suporte de armazenamento, e as datas em que o refrescamento foi feito pela última vez), (iii) a obsolescência de formatos deve ser combatida com estratégias que garantam que os recursos possam continuar a ser utilizados, seja por migração ou por emulação (tanto uma como outra requerem metadados sobre os formatos de origem e o ambiente tecnológico que os suporta) e (iv) as estratégias de preservação podem provocar alterações nos objectos digitais originais ou na forma como estes são reproduzidos, o que pode atentar contra a autenticidade do objecto (os metadados podem ajudar a suportar a autenticidade de um recurso mediante a documentação da proveniência digital de tal objecto – a sua cadeia de custódia e o historial de alterações autorizadas).

No contexto da preservação digital, os metadados ganham especial importância porque, segundo Ferreira (2006, p. 54): (i) procuram descrever o caminho percorrido pelos documentos digitais, desde a sua criação até à sua incorporação no repositório, (ii) descrevem o ambiente tecnológico que sustenta a correcta apresentação dos objectos digitais (*hardware, software* e sistemas operativos), (iii) assumem também a responsabilidade de garantir a sua autenticidade e (iv) incluem informação sobre todas as contingências legais que afectam os objectos digitais

⁸³ Para que possam ser usados no seu potencial máximo, os metadados devem ser representados através de esquemas, ou seja, estruturas de dados normalizadas e controladas. Um esquema de metadados corresponde a um conjunto de elementos que visa ir ao encontro das necessidades de comunidades particulares. Alguns esquemas são de natureza geral, mas a maioria é criada para tipos específicos de informação. Os esquemas de metadados podem variar grandemente, quer no número de dados dos elementos, quer no uso de elementos principais ou repetíveis, quer na forma como são codificados, quer no uso de vocabulários controlados. Alguns esquemas de metadados concentram-se nos elementos descritivos que suportam a pesquisa e a recuperação da informação (*Dublin Core*, por exemplo); outros esquemas estão mais virados para os elementos administrativos e/ou estruturais, dependendo sempre do seu objectivo e aplicação.

custodiados pelo repositório. Os metadados são ainda capitais na medida em que documentam as acções e estratégias de preservação levadas a cabo ao longo do tempo e registam as consequências que essas acções tiveram sobre a aparência e funcionalidade dos objectos digitais. Em suma, os metadados de preservação são de extrema importância num repositório digital, porque permitem que os objectos digitais custodiados permaneçam auto-documentados ao longo do tempo.

5. Aspectos legais da preservação digital

No âmbito da preservação digital, o depósito legal e os direitos de autor são aspectos legais cruciais, uma vez que a aplicação de grande parte das estratégias de preservação depende da forma como estes trâmites legais são executados. Depósito legal e direitos de autor são conceitos com que as bibliotecas estão familiarizadas. Referi-los no contexto da preservação digital faz todo o sentido, pois as inovações tecnológicas que estão por detrás da forma como os documentos têm vindo a ser publicados, distribuídos, disseminados e preservados têm motivado uma longa reflexão e um aceso debate com o objectivo de adaptar estes dispositivos legais ao novo contexto tecnológico e aos imperativos da era digital.

Por um lado, até à proliferação dos documentos digitais, o depósito legal permitiu que determinadas bibliotecas colectassem, disponibilizassem e preservassem todas as publicações editadas num determinado país. Por outro lado, os direitos de autor (que, regra geral, incluem excepções e limitações em benefício das bibliotecas e instituições similares), possibilitaram que elas fizessem um uso mais justo dos documentos que integram as suas colecções, concedendo-lhes os trâmites legais para que levassem a cabo a sua missão de divulgação e preservação, sem violação da lei de direitos.

Perante a necessidade urgente de adaptar estes dispositivos legais ao contexto digital e às suas especificidades, é preciso considerar o depósito legal e os direitos de autor como mecanismos importantes de colecta e preservação do todo o património literário por parte das instituições que têm assumido essa responsabilidade.

A impossibilidade de colectar todos os documentos digitais através do depósito legal e o facto de os direitos de autor terem vindo a ser reforçados no mundo digital continuam a ser os maiores obstáculos à preservação digital. A maioria das estratégias de preservação digital pressupõem várias cópias de um documento e/ou a manipulação e alteração da sua apresentação e funcionalidade, o que pressupõe o exercício de direitos exclusivos do autor, incluindo o direito de reprodução e o direito de assegurar a genuinidade e integridade da sua obra:

Digital preservation strategies in use or proposed involve copying and possibly changing information. This raises complex rights questions. Copying may have to be repeated periodically. The rights in content and any associated software may belong to a number of different individuals or organizations. Digital information may be surrounded by technological protection preventing unauthorized copying and redistribution that may also inhibit or prevent preservation actions. (Muir, 2004, p. 73)

A execução destas actividades sem autorização dos detentores dos direitos de autor pode resultar numa infracção, a menos que: (i) os documentos se encontrem em domínio público; (ii) seja o próprio detentor dos direitos de autor a levar a cabo as actividades de

preservação; (iii) o detentor dos direitos de autor autorize a execução de tais actividades; (iv) se apliquem as excepções previstas nas leis sobre os direitos de autor aplicadas, por exemplo, às bibliotecas.

A continuidade do património digital e o sucesso da sua preservação dependem do acesso a direitos básicos necessários às actividades de preservação: (i) o direito de obter os documentos digitais; (ii) o direito de fazer cópias para fins de preservação; (iii) o direito de contornar os dispositivos/mecanismos de protecção utilizados pelos detentores dos direitos para impedir a cópia dos documentos; (iv) o direito de adicionar metadados ao objecto digital; (v) o direito de modificar as estruturas e nomes dos ficheiros, se necessário; e, cumprindo o objectivo primordial da preservação, (vi) o direito de tornar os documentos preservados acessíveis.

No entanto, as bibliotecas têm enfrentado algumas dificuldades na obtenção destes direitos seja porque muitas vezes os seus detentores não autorizam o direito de reprodução ou de alteração da apresentação, seja pela natureza fragmentária ou colectiva dos documentos digitais que pode tornar difícil a localização todos os detentores dos direitos de forma a negociar com todos eles as autorizações necessárias, seja mesmo porque a deslocalização inerente aos documentos publicados na *Web*, pode dificultar a determinação da jurisdição legal competente que pode ser diferente de país para país (Webb, 2003, p 102).

5.1. Mecanismos de protecção como entraves à preservação digital

Perante a maleabilidade característica dos objectos digitais e a facilidade com que estes podem ser duplicados e distribuídos, autores e distribuidores recorrem a mecanismos que fortalecem a sua protecção. Num esforço de verem os seus direitos reforçados, muitos autores⁸⁴ optaram por confiar a divulgação dos seus trabalhos a editoras, as quais, através de contratos de licenciamento combinados com outras medidas de protecção tecnológica, regulam o acesso e uso dos documentos. A combinação destes dois tipos de protecção pode conduzir em extremo a uma protecção exacerbada dos interesses dos detentores do direito de autor em detrimento do interesse público, deixando cair por terra as excepções e limitações ao direito de autor conferidas às bibliotecas.

⁸⁴ Ao abordar a questão dos direitos de autor, devemos distinguir dois tipos de autores: aqueles cuja produção representa o seu meio de subsistência (autores de obras literárias, guionistas, cronistas,...) e aqueles cuja produção apenas serve para promoção intelectual (investigadores, académicos,...) que normalmente cedem os seus direitos a editores não esperando outro rendimento que não seja a divulgação e reconhecimento do seu trabalho.

Segundo Koelman & Helberger (2000, apud Fernández-Molina, 2007) é possível distinguir quatro tipos de sistemas de protecção tecnológica: os que controlam o acesso (mediante a utilização de uma palavra-chave); os que controlam certos usos (impedindo a cópia); os que protegem a integridade do documento (marcas de água digitais, encriptação, etc.); os que exigem o pagamento para acesso ou uso da informação (contabilizam o seu uso, detectam infracções ou incumprimentos do licenciamento). Verifica-se, pois que a protecção dos documentos digitais pode ser de dois tipos: contratual ou tecnológica.

5.1.1. Protecção contratual

Este tipo de protecção efectua-se através de contratos de licenciamento, em que as editoras e/ou distribuidoras negociam com as bibliotecas e instituições similares licenças de utilização dos conteúdos depositados nas suas bases de dados.

AS BASES DE DADOS E OS CONTRATOS DE LICENCIAMENTO

Em Portugal, o decreto-lei⁸⁵ que regula a protecção jurídica das bases de dados considera estas como “criações intelectuais” e, por isso, “protegidas pelo direito de autor com algumas especificidades”. Assim, aos produtores de certas bases de dados⁸⁶ é concedida uma protecção *sui generis* com a duração de 70 anos. Este direito *sui generis* confere ao fabricante da base de dados o direito de autorizar ou proibir a extracção (a transferência, permanente ou temporária, da totalidade ou de uma parte substancial do conteúdo de uma base de dados para outro suporte) e/ou a reutilização (qualquer forma de distribuição ao público da totalidade ou de uma parte substancial do conteúdo da base de dados, nomeadamente através da distribuição de cópias, aluguer, transmissão em linha ou outra modalidade) da totalidade ou de uma parte substancial do seu conteúdo. Quer isto dizer que quem não for utilizador legítimo da base de dados, ou seja, quem não tiver celebrado um contrato de acesso aos objectos digitais aí depositados, não pode proceder à sua extracção ou reutilização. Assim, as bibliotecas apenas podem aceder, extrair e reutilizar o conteúdo de uma base de dados se existir um contrato de licenciamento. O referido Decreto-lei confere-lhes as mesmas excepções previstas no artigo 75º do *Código de Direito de Autor e Direitos Conexos*, mas só e apenas se existir um contrato de

⁸⁵ Decreto-Lei nº 122/2000 de 4 de Julho

⁸⁶ Segundo o artigo 12º, nº1 desse Decreto-Lei, uma base de dados está protegida pelos direitos de autor “quando a obtenção, verificação ou apresentação do conteúdo de uma base de dados represente um investimento substancial do ponto de vista qualitativo ou quantitativo (...)”.

licenciamento. Evidentemente, para uma biblioteca, ainda que nacional, é economicamente inviável – já que o acesso a essas bases de dados é, regra geral, pago e, normalmente, é feito por autenticação – estabelecer contratos de licenciamento com todas as bases de dados nacionais e/ou internacionais. Com o Decreto-Lei nº 122/2000 de 4 de Julho caem por terra os desígnios da Lei de Depósito Legal⁸⁷, que tem como objectivo a criação de uma colecção de toda a produção nacional, bem como a sua divulgação e preservação.

Mesmo tendo acesso a uma base de dados, a protecção contratual estabelecida pelo fabricante da base de dados pode proporcionar apenas uma transferência limitada de direitos sobre um determinado documento. Assim, e contrariamente ao que acontecia com os documentos analógicos (que chegavam às bibliotecas por meio de compra, doação ou depósito legal), quando recorrem ao licenciamento as bibliotecas não recebem, não armazenam, nem tão-pouco possuem o documento digital, mas apenas detêm o direito de acesso e, também esse, em muitos casos, limitado (Fernández-Molina, 2007). No caso dos recursos em continuação (por exemplo, revistas científicas), as bibliotecas possuíam os números que iam sendo publicados e, caso a sua publicação cessasse, tinham sempre a garantia de posse dos números que receberam, e estes permaneceriam disponíveis para os utilizadores e seria possível preservá-los para um futuro acesso. Tal pode não acontecer quando existe um contrato de licenciamento em que se negocia apenas o acesso. No caso de uma publicação cessar, o editor pode cessar também o acesso aos números anteriores⁸⁸.

PRODUTORES DE SOFTWARE E LICENÇAS DE UTILIZAÇÃO

Outro aspecto a ter em conta quando se consideram os aspectos legais da preservação da informação digital é o facto de a aplicação das estratégias de preservação de objectos digitais implicar a obtenção de licenças de utilização da tecnologia requerida para interpretar os objectos digitais, sobretudo daquelas associadas às aplicações de *software* e aos formatos. A menos que se utilizem aplicações e formatos abertos, a maioria dos objectos digitais, na sua dimensão lógica, dependem de licenças de utilização. Quando proprietárias, essas licenças têm um custo para as bibliotecas que ficam dependentes de um fornecedor (por exemplo, a Microsoft) que, no futuro, pode deixar de assegurar o funcionamento dessas aplicações e/ou formatos.

⁸⁷ Ver 5.2. *Depósito legal dos documentos digitais*.

⁸⁸ Hoorens et al. (2007, p. 7) defendem que as bibliotecas devem ficar na posse dos documentos digitais cujo acesso foi negociado através de contratos de licenciamento, pois entendem que existem factores que podem comprometer o acesso perpétuo a esses documentos.

Em muitos casos, o acesso aos objectos digitais depende de licenças de utilização que protegem as aplicações de *software* e os formatos que interpretam os documentos digitais e os tornam legíveis. Devido a constrangimentos financeiros, as bibliotecas ficam impedidas de adquirir todas as licenças necessárias para aceder e preservar os vários documentos digitais sob a sua custódia. Deste modo, também os produtores de *hardware* e *software* devem assumir um papel de relevância neste domínio, conferindo às bibliotecas que levam a cabo projectos de preservação digital excepções de utilização dos seus produtos.

Os direitos limitados impostos pelos contratos de licenciamento e pelas licenças de utilização que protegem as aplicações de *software* e os formatos não estão em consonância com os privilégios conferidos pelo direito de autor às bibliotecas (direito de reprodução, por exemplo) e impossibilitam a sua missão de preservação. Sem disporem de uma cópia física dos documentos e sem possuírem as aplicações de *software* necessárias para a sua interpretação, as bibliotecas não podem aplicar as estratégias de preservação digital necessárias. Uma solução para este problema passará pela negociação com as editoras com as quais as bibliotecas estabelecem acordos de licenciamento e com os produtores de *software*: negociação os direitos que essas entidades deterão sobre os documentos cujo acesso já foi pago, para que as bibliotecas assegurem a preservação dos documentos.

5.1.2. Protecção tecnológica

Os autores e editores têm recorrido a mecanismos tecnológicos para protegerem os documentos digitais de usos indevidos e ilegais. Trata-se de dispositivos que protegem tecnologicamente os documentos tutelados pelo direito de autor, controlando o seu acesso e o seu uso e conferindo-lhes autenticidade. A criptografia⁸⁹, a esteganografia⁹⁰ e as marcas de água digitais são alguns exemplos de protecção tecnológica.

A protecção tecnológica é ainda mais eficaz, porque é permitida por lei na maior parte dos casos: qualquer tentativa de iludir os mecanismos de protecção tecnológica constitui uma infracção à lei.

⁸⁹ Consiste na protecção de uma mensagem, codificando o texto em linguagem natural num texto cifrado. O acesso faz-se através de chaves secretas que devem manter-se secretas com o objectivo de proteger a informação. Em ambiente digital onde os dados são representados por *bits*, o processo de criptografia é feito através de algoritmos.

⁹⁰ Técnica semelhante à criptografia que, em vez de ocultar o significado da mensagem, oculta a existência da mensagem.

Ora, para levarem a cabo as suas estratégias de preservação digital as bibliotecas terão, necessariamente (em ambiente digital, preservar significa copiar), de contornar essas medidas de protecção sem que incorram na violação da lei. E isto representa um problema grave na preservação da informação:

Este punto puede llegar a representar un grave riesgo de cara a la preservación, ya que si no se puede utilizar legalmente mecanismos para suprimir las medidas tecnológicas de protección, será imposible, a largo plazo, acceder a muchas obras, porque habrán desaparecido los códigos de paso de la protección o del organismo o el mecanismo de autorización, o simplemente por la imposibilidad de hacer una copia de refresco debido a que ésta saldrá con una degradación sustancial respecto al original a causa de las medidas de protección integradas. (Keefler & Gallart, 2000, p. 169)

Aludiu-se ao facto dos documentos digitais se apresentarem, regra geral, protegidos por mecanismos tecnológicos de protecção que impedem a cópia e distribuição. Mas há ainda outra questão. Na verdade, mesmo que os titulares do direito de autor permitam que tais mecanismos sejam contornados pelas instituições beneficiárias das excepções e limitações ao direito de autor, os proprietários do *software* utilizado podem impedir que se levem a cabo algumas estratégias de preservação digital.

5.2. Depósito legal dos documentos digitais

O depósito legal é uma obrigação estabelecida por lei, que obriga todas as entidades, públicas ou privadas, e todas as pessoas que produzam qualquer tipo de documento em múltiplos exemplares a depositar um ou mais exemplares numa instituição nacional reconhecida com vista à criação de uma colecção nacional que deve ser preservada.

Em Portugal, o decreto-lei que regula o depósito legal data de 1982 e não sofreu qualquer alteração até à data. No entanto, nestes quase 30 anos, alterou-se profundamente o modo de publicação e distribuição de documentos. Grande parte da informação começou a ser publicada em formato digital e passou a ser distribuída na *Web*, pelo que se torna difícil colectar e identificar a origem dessa informação. Deste modo, é imperativo que o legislador actualize este dispositivo legal para que os seus objectivos se cumpram também na era digital:

Traditional preservation methods, such as the 'legal deposit' used by national libraries to ensure that copies of all printed materials are kept, cannot be applied in the same way to digital material for a variety of reasons, notably because web publications often draw on data stored on servers in different parts of the world. The sheer volume of data involved also poses a problem. (Abid, 2007, p. 9)

Na tentativa de ajudar os países na elaboração, actualização e revisão das suas legislações sobre o depósito legal, a UNESCO, em parceria com a IFLA – e sob a direcção de

Jules Larivière – actualizou, em 2000, o documento *Guidelines for legal deposit legislation*, tendo em conta o novo contexto tecnológico. Apesar de alguns países já terem alargado o depósito legal aos documentos digitais (Suécia, Dinamarca, África do Sul,...), em Portugal continua em vigor uma lei desactualizada que compromete a missão das bibliotecas depositárias: garantir o acesso universal e equitativo à informação e preservar todas as publicações editadas no País independentemente do seu formato⁹¹.

A lei portuguesa é bastante clara quanto aos objectivos do depósito legal: (i) defender e preservar os valores da língua e cultura portuguesas; (ii) constituir e conservar uma colecção nacional (todas as publicações editadas no País); (iii) produzir e divulgar a bibliografia nacional corrente; (iv) estabelecer a estatística das edições nacionais; (v) enriquecer as bibliotecas dos principais centros culturais do País. No entanto, para que tais objectivos se cumpram há que adaptar a legislação ao contexto actual, alargando o tipo e formato das obras objecto de depósito legal⁹² e contemplando o facto de a informação publicada na *Web* ser maioritariamente deslocalizada⁹³.

A não contemplação dos documentos digitais nos tipos de documentos objecto de depósito legal pode comprometer (se é que já não comprometeu) seriamente os propósitos de criação de uma colecção nacional e de divulgação e preservação dessa mesma colecção. O texto de um projecto de Decreto-lei relativo ao Depósito Legal foi distribuído pelo Secretário de Estado da Cultura em reunião do Conselho Superior de Bibliotecas, ocorrida em Setembro de 2002 com o objectivo de regulamentar e adequar a lei do depósito legal. Nesse texto, é evidente a vontade de alargar o depósito legal às obras em formatos electrónicos: no entanto, que essa obrigatoriedade apenas recairia sobre as todas as publicações electrónicas com suportes físicos. Tratando-se apenas de uma proposta, esse texto não abrange as publicações electrónicas sem suporte físico, tais como as difundidas ou acessíveis em redes de comunicação de dados como a

⁹¹ A iniciativa da BN *DiTeD - Teses e Dissertações* que se destina a gerir e a preservar as teses e dissertações portuguesas em formato digital, apesar de notável, permanece insuficiente porque destina-se apenas a um tipo de documentos específico e porque o seu depósito é voluntário.

⁹² Segundo o Decreto-lei nº 74/82, “são objecto de depósito legal as obras impressas ou publicadas em qualquer ponto do País, seja qual for a sua natureza e o seu sistema de reprodução, (...)” (Portugal, 1982), sendo obrigatório o depósito de livros, brochuras, revistas, jornais e outras publicações periódicas, separatas, atlas e cartas geográficas, ... selos, estampas,... fonogramas e videogramas, obras cinematográficas, microformas e outras reproduções fotográficas. O Decreto-lei nº 362/86 de 28 de Outubro, estende a obrigatoriedade de depósito legal às teses de doutoramento e mestrado, às dissertações e às provas de aptidão. É fundamental clarificar o que se entende por “natureza e o seu sistema de reprodução” para que as bibliotecas saibam se podem incluir os documentos digitais no tipo de obras objecto de depósito legal.

⁹³ Se o objectivo é construir uma colecção nacional com todas as publicações editadas no país, então a nova legislação deve contemplar o facto de muitas vezes ser difícil localizar e identificar a proveniência dos documentos que se encontram *online*.

Internet, que seriam objecto de depósito legal selectivo, nos termos dos acordos a estabelecer entre a BN e os respectivos produtores⁹⁴.

Aquele diploma mereceu uma tomada de posição⁹⁵ da *Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas* (APBAD), a qual, congratulando-se pelo facto de a nova legislação considerar “a abstracção do conceito de publicação relativamente aos seus suportes”, salienta que o diploma, não abrange todas as publicações electrónicas, independentemente de possuírem ou não suporte⁹⁶. Volvidos oito anos desde a publicação do diploma, nenhum outro documento legal sobre o assunto foi publicado.

Pese embora esta necessidade de alargamento do depósito legal aos documentos digitais, a verdade é que a facilidade em estabelecer no mundo impresso o que se considera publicação sujeita a depósito legal não se verifica no mundo digital: a distinção feita no diploma entre publicações electrónicas com suporte físico e sem suporte físico é pertinente e, por isso, susceptível de ser discutida.

Segundo Larivière (2000, p. 31), quando se advoga a extensão do depósito legal aos documentos digitais, devem ser tidas em conta duas categorias deste tipo de documentos: os *documentos digitais offline* e os *documentos digitais online*. A contemplação e controlo dos primeiros na legislação de depósito legal é relativamente simples: sendo estes publicados e divulgados num suporte tangível (como CD-ROM ou DVD), podem ser adaptados os procedimentos usados em relação aos documentos impressos. O depósito de documentos digitais em suporte físico é prática corrente em muitos países (Borbinha, 2004). Dado que os documentos digitais requerem de tecnologias que possibilitem a sua leitura, Larivière (2000, p. 31) sugere que novas legislações sobre depósito legal, não só devem considerar os documentos digitais como objecto de depósito, mas também devem exigir aos depositantes o depósito de manuais e de outros materiais necessários para a utilização e leitura desses documentos.

⁹⁴ Borbinha (2004) enumera algumas formas de a BN poder entrar na posse física de publicações digitais que incluem a colecta de recursos publicados na *Web* através de sistemas de recolha automatizados e o depósito que pode ser voluntário ou contratual.

⁹⁵ Disponível na WWW: <http://www.apbad.pt/downloads/tpDeposito.pdf>

⁹⁶ Na sua tomada de posição, a APBAD prefere destacar, e bem, o facto de a nova legislação pretender reduzir o número de exemplares a depositar (de 14 para 2) e querer centralizar o depósito legal numa única instituição, a Biblioteca Nacional. Se tal se concretizar, a preservação e acessibilidade do património nacional, analógico ou digital, estarão seriamente comprometidas. Segundo a APBAD (2002), “a fixação em dois do número de exemplares a depositar (...), carece de fundamentação e põe em causa a preservação e a acessibilidade em todo o território das publicações nacionais. (...) É contraproducente a centralização do Depósito Legal numa única instituição, com tudo o que de grave e inconveniente poderá daí resultar, nomeadamente em situações de catástrofe, acidente ou terrorismo.”

Naturalmente, tal como acontece com os documentos impressos, qualquer actualização ou nova versão do documento digital constitui um novo objecto de depósito legal.

No que concerne aos *documentos digitais online*, a contemplação e controlo do depósito legal já não se revela tão simples. Contrariamente aos *documentos digitais offline*, em que existem múltiplos exemplares que podem ser distribuídos pelas bibliotecas depositárias, dos *documentos digitais online*, regra geral, existe apenas uma cópia que está armazenada em qualquer computador central onde pode ser actualizada constantemente, sob a posse de um editor ou distribuidor (Larivière 2000, p. 32).

Esta deslocalização e intangibilidade dos *documentos digitais online* coloca problemas às bibliotecas, pois o acesso a esses documentos, ou é controlado pela exigência de uma palavra-chave de acesso, ou está sujeito a pagamento. Muitas bibliotecas, para poderem ter acesso a *documentos digitais online*, efectuam contratos de licenciamento. Este tipo de licença confere o acesso, mas não cumpre os objectivos do depósito legal, visto que as bibliotecas não possuem uma cópia efectiva e tangível dos documentos que lhes possibilite a construção duma colecção a preservar e divulgar.

5.3. Direitos de autor e a preservação digital: o caso português

O direito de autor e direitos conexos das obras literárias e artísticas são direitos declarados internacionalmente através de convenções e tratados: desde a mais antiga convenção internacional sobre o copyright – a *Convenção de Berna* (1886) – até aos diversos tratados que têm vindo a responder às necessidades da era digital: *Trade-Related Aspects of Intellectual Property Rights (TRIP)*⁹⁷, de 1994; *WIPO Copyright Treaty (WCT)*⁹⁸, de 1996; *Directive 2001/29/EC of the European Parliament and of the Council*⁹⁹, de 2001.

Segundo o *Código do Direito de Autor e dos Direitos Conexos* (Portugal, 2008, p. 1902) o direito de autor abrange “direitos de carácter patrimonial”, isto é, o direito exclusivo do autor dispor da sua obra e de fruí-la e utilizá-la, ou autorizar a sua fruição ou utilização por terceiro, total ou parcialmente, e os chamados “direitos morais”, ou seja, o direito do autor reivindicar a paternidade de uma obra e assegurar a sua genuidade e integridade. Este direito aplica-se a todas as criações intelectuais do domínio literário, científico e artístico, independentemente do tipo de

⁹⁷ Disponível na WWW: http://www.wto.org/english/tratop_e/trips_e/t_agm0_e.htm

⁹⁸ Disponível na WWW: http://www.wipo.int/export/sites/www/treaties/en/ip/wct/pdf/trtdocs_wo033.pdf

⁹⁹ Também conhecida como a *Directiva da Sociedade da Informação*

suporte utilizado para a sua exteriorização. Na legislação portuguesa, o direito moral é um direito inalienável, irrenunciável e que se perpetua após a morte do autor, enquanto o direito patrimonial é um direito transmissível que caduca 70 anos¹⁰⁰ após a morte do autor, altura em que a obra entrará em domínio público.

Como vem sendo estipulado desde a *Convenção de Berna*¹⁰¹, o direito de autor não deve ser entendido um direito total, absoluto e discricionário. Ele é também um direito que visa protecção das obras e a sua divulgação. Por essa razão, as várias legislações sobre o direito de autor prevêm limitações e excepções baseadas em razões de interesse público, como a comunicação e a divulgação das obras e a promoção da educação, da investigação e da cultura. A utilização livre é uma das excepções ao direito de autor, em benefício das bibliotecas e instituições similares, conferindo-lhes o direito de reprodução por razões de investigação e ensino, arquivo e/ou preservação.

Estas excepções e limitações ao direito de autor em benefício das bibliotecas não só visam uma utilização mais justa das obras, tornando-as acessíveis a um maior número de pessoas, como também permite que as bibliotecas cumpram a sua missão enquanto instituições de preservação do património literário, cultural e científico.

Durante décadas, as bibliotecas usufruíram dessas excepções ao direito de autor, a fim de prosseguirem a sua missão de preservação das suas colecções, ora utilizando mecanismos de foto-reprodução (a microfilmagem, por exemplo), ora, mais recentemente, procedendo à digitalização de documentos, com especial incidência naqueles que se encontram em domínio público. Até à proliferação dos documentos digitais, a aplicação das excepções e limitações do direito de autor para fins de preservação foi sendo pacífica e não levantava grandes questões quanto ao incumprimento ou infracção a esse direito. No entanto, as características específicas dos documentos digitais e a forma como estes são distribuídos levantam questões legais que normalmente não se colocavam aquando da preservação dos documentos analógicos.

As ameaças à integridade e autenticidade dos documentos digitais e a sua fácil distribuição e duplicação levaram os detentores dos direitos de autor (os próprios autores ou os editores/distribuidores) a utilizar mecanismos de protecção (como foi referidos, esses mecanismos visam, sobretudo, impedir a cópia e distribuição dos documentos). Ora, em ambiente digital, preservar implica fazer cópia do documento original. Todas as estratégias de

¹⁰⁰ A duração do direito de autor ultrapassa em muito o tempo de intervenção que permite às bibliotecas aplicarem estratégias de preservação com sucesso. Por isso, não é viável esperar que as obras entrem em domínio público para que possam aplicar-se as estratégias de preservação necessárias.

¹⁰¹ Nos artigos 9º, 10º, 10º bis, 11º bis prevê-se que seja permitida a reprodução de uma obra (se tal não prejudicar os interesses económicos ou outros interesses legítimos do autor), a citação de partes da obra, a reprodução de artigos de publicações periódicas e a partilha e comunicação das obras respectivamente.

preservação digital colidem, de uma maneira ou de outra, com o direito de autor, se as exceções reservadas às bibliotecas não forem tidas em conta. O sucesso da preservação digital depende da regulamentação e adaptação das exceções e limitações ao direito de autor em ambiente digital.

Com o intuito de harmonizar o direito de autor com o direito de preservação (que as bibliotecas e instituições congêneres têm assumido desde há várias décadas), os esforços nos últimos anos vão no sentido de reapreciar, à luz do novo ambiente tecnológico, as exceções e limitações ao direito de autor promulgadas pela *Convenção de Berna*.

O *Tratado da Organização Mundial da Propriedade Intelectual sobre o Direito de Autor* (1996) constituiu um importante ponto de partida para a reforma das legislações sobre o direito de autor dos países subscritores de forma a adaptá-las ao contexto digital.

Em consonância com a *Convenção de Berna*, o artigo 10º daquele tratado reitera as exceções e limitações ao direito de autor, prevendo usos excepcionais das obras: por exemplo, a sua reprodução para preservação. A originalidade relativamente à convenção de 1886 reside numa nota de rodapé onde é feito apelo aos países signatários para que estendam estas limitações e exceções aos documentos digitais¹⁰², normalmente protegidos por mecanismos tecnológicos de protecção que impedem as bibliotecas de aplicar as estratégias de preservação digital. Apesar de este tipo de protecção ser assegurado no artigo 11º, o *Tratado da Organização Mundial da Propriedade Intelectual sobre o Direito de Autor*, onde se impede e pune a ilusão dessas medidas de protecção, também recomenda que se abram as devidas exceções que permitam às bibliotecas cumprirem a sua missão de preservação¹⁰³. Embora se trate de meras recomendações, os artigos 10º e 11º daquele tratado constituem a primeira tentativa, ainda que de forma tímida e pouco esclarecedora, de adequação das exceções e limitações ao direito de autor em benefício das bibliotecas (no que concerne à preservação dos documentos digitais).

¹⁰² “It is understood that the provisions of Article 10 permit Contracting Parties to carry forward and appropriately extend into the digital environment limitations and exceptions in their national laws which have been considered acceptable under the Berne Convention. Similarly, these provisions should be understood to permit Contracting Parties to devise new exceptions and limitations that are appropriate in the digital network environment” (WIPO, 1996, p.4).

¹⁰³ “Contracting Parties shall provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures that are used by authors in connection with the exercise of their rights under this Treaty or the Berne Convention and that restrict acts, in respect of their works, which are not authorized by the authors concerned or **permitted by law**” (WIPO, 1996, p.4). Subentende-se que as acções são permitidas por lei quando contempladas como exceções aos direitos de autor. Nessas acções poderemos incluir o direito de cópia para preservação.

No virar de século, a necessidade de fomentar um justo equilíbrio entre o direito de autor e o interesse público no seio da Sociedade da Informação é reforçada pela *Directiva 2001/29/EC* da União Europeia (UE) que esteve na origem da reforma dos códigos de direito de autor dos 27 países membros. A directiva não é clara no que concerne às excepções e limitações ao direito de autor que permitem às bibliotecas e instituições similares levar a cabo a sua missão de preservação. Na alínea c) do nº 2 do artigo 5º, a directiva refere que os “Estados-Membros podem prever excepções ou limitações ao direito de reprodução (...) em relação a actos específicos de reprodução praticados por bibliotecas, estabelecimentos de ensino ou museus acessíveis ao público, ou por arquivos, que não tenham por objectivo a obtenção de uma vantagem económica ou comercial, directa ou indirecta” (UE, 2001, p. 16). No entanto, não se especifica que tipo de cópia pode ser feita (analógica ou digital) ou que tipo de material pode ser reproduzido, nem se clarificam quais os objectivos das cópias resultantes da aplicação destas excepções. Esta ausência de transparência pode, segundo Fernández-Molina (2007), permitir que se inclua nestas excepções o direito à cópia com fins de preservação, mas tal efeito dependerá sempre da forma como cada país membro fizer a transposição do artigo 5º, o qual, como foi adiantado, é de carácter facultativo.

Ainda relacionado com as cópias para preservação está o artigo 6º onde se definem as obrigações relativamente às medidas de carácter tecnológico. Segundo este artigo, os Estados-Membros têm a obrigação de criar mecanismos de protecção jurídica contra o contorno dos mecanismos tecnológicos de protecção utilizados pelos autores para salvaguardarem os seus direitos. Contudo, o nº 4 deste mesmo artigo adianta que, apesar de esta protecção jurídica, “na falta de medidas voluntárias tomadas pelos titulares de direitos, nomeadamente de acordos entre titulares de direitos e outras partes interessadas, os Estados-Membros tomarão as medidas adequadas para assegurar que os titulares de direitos coloquem à disposição dos beneficiários de excepções ou limitações previstas na legislação nacional (...) os meios que lhes permitam beneficiar dessa excepção ou limitação, sempre que os beneficiários em questão tenham legalmente acesso à obra ou a outro material protegido em causa” (UE, 2001, p.17-18). À semelhança do que acontece com o *Tratado da Organização Mundial da Propriedade Intelectual sobre o Direito de Autor*, a directiva da UE prevê que um Estado-Membro tome medidas de protecção relativamente a um beneficiário de uma excepção ou limitação prevista pelo código de direito de autor e em conformidade com artigo 5º da mesma directiva. Assim, aquando da transposição, o Estado-Membro pode permitir o contorno de medidas de protecção tecnológica por parte das bibliotecas, sempre que tal se justifique para fins de preservação, por exemplo.

Sendo a adopção das excepções e limitações veiculadas no artigo 5º de carácter facultativo e estando o nº 4 do artigo 6º dependente dessa adopção, o sucesso da preservação

digital nos países da UE depende da forma como os diferentes países efectuarem a transposição da directiva. Segundo o estudo sobre as limitações e excepções ao direito de autor em benefício das bibliotecas e dos arquivos, levado a cabo por Kenneth Crew (2008), apenas 21 dos 27 países da UE aprovaram pelo menos uma das excepções do artigo 5º da directiva, podendo ela incluir ou não o direito de as bibliotecas efectuarem cópias para fins de preservação.

A transposição da *Directiva 2001/29/EC* da UE para legislação portuguesa foi consolidada com publicação da Lei nº 16/2008 que tem como anexo o *Código de Direito de Autor e Direitos Conexos*. Portugal é um dos 21 países que incluem no seu código excepções e limitações ao direito de autor ao abrigo do artigo 5º da directiva da UE. Na alínea e) do artigo 75º do Capítulo II intitulado “Da utilização livre” o código português estipula que:

“a reprodução, no todo ou em parte, de uma obra que tenha sido previamente tornada acessível ao público [é lícita], desde que tal reprodução seja realizada por uma biblioteca pública, um arquivo público, um museu público, um centro de documentação não comercial ou uma instituição científica ou de ensino, e que essa reprodução e respectivo número de exemplares se não destinem ao público, se limitem às necessidades das actividades próprias destas instituições e que não tenham por objectivo a obtenção de uma vantagem económica ou comercial, directa ou indirecta, incluindo os actos de reprodução necessários à preservação e arquivo de quaisquer obras” (Portugal, 2008, p. 1910).

O referido artigo afigura-se mais objectivo do que o artigo 5º da directiva: nele referem-se que tipo de obras podem ser copiadas e com que objectivos. No que concerne à preservação digital, é de salientar o facto de a lei portuguesa permitir que se façam cópias (não especificando o número) necessárias à “preservação e arquivo de quaisquer obras”¹⁰⁴.

No que se refere à ilusão dos mecanismos de protecção tecnológica com o objectivo de efectuar cópias de preservação, a lei portuguesa, indo ao encontro daquilo que foi estipulado pelo *Tratado da Organização Mundial da Propriedade Intelectual sobre o Direito de Autor* e pela *Directiva 2001/29/EC* da EU, permite que as bibliotecas a executem, não considerando que se trate, nestes casos, de uma infracção à lei. Se, por um lado, o artigo 217º do *Código de Direito de Autor e Direitos Conexos* português protege os titulares dos direitos de autor e os titulares do direito *sui generis* contra a neutralização de medidas de carácter tecnológico¹⁰⁵ utilizadas para a protecção das suas obras por parte de terceiros, por outro lado, são também

¹⁰⁴ No código, consideram-se obras as criações intelectuais do domínio literário, científico e artístico, por qualquer modo exteriorizadas.

¹⁰⁵ “Entende-se por «medidas de carácter tecnológico» toda a técnica, dispositivo ou componente que, no decurso do seu funcionamento normal, se destinem a impedir ou restringir actos relativos a obras, prestações e produções protegidas, que não sejam autorizados pelo titular dos direitos de propriedade intelectual” (Portugal, 2008, p. 1930)

estipuladas, no artigo 221º, limitações à protecção dessas medidas tecnológicas, sendo também apresentados os mecanismos legais de que as bibliotecas dispõem para beneficiar deste direito¹⁰⁶.

Da análise dos artigos 75º e 221º do *Código de Direito de Autor e Direitos Conexos* português pode concluir-se, portanto, que as excepções e limitações ao direito de autor em benefício das bibliotecas estão, em princípio, salvaguardadas no contexto digital, nomeadamente o direito de as bibliotecas efectuarem cópias de preservação, ainda que tal não seja estipulado com a clareza necessária. Naturalmente, sendo essa acção permitida, o código prevê também que os detentores do direito de autor não utilizem medidas tecnológicas de protecção que impeçam as bibliotecas de efectuar tais cópias e, se tal acontecer, o código obriga os detentores dos direitos a depositar junto da Inspecção-Geral de Actividades Culturais (IGAC) os meios que permitem às bibliotecas beneficiar das formas de utilização legalmente permitidas.

Falta esclarecer, no entanto, quando e em que circunstâncias podem as bibliotecas efectuar as cópias de preservação. No caso dos documentos digitais, podem essas cópias ser feitas sempre que se verifique algum tipo de obsolescência tecnológica? E quantas cópias são permitidas? As cópias terão de ser cópias exactas do original ou pode haver perda de informação e a integridade ser afectada? O direito à cópia para preservação é extensível a todos os documentos, incluindo aqueles que estão sob a tutela de editoras que os distribuem através de contratos de licenciamento? A resposta a estas questões no contexto da preservação digital é importante pois auxilia as bibliotecas no momento da aplicação das diferentes estratégias de preservação.

Quando se trata da preservação de documentos digitais, a coordenada tempo é um factor crucial: convém que se estabeleça um intervalo de tempo mínimo de acção sob pena de ser demasiado tarde para a aplicação eficaz de uma qualquer estratégia de preservação. Além disso, estratégias como o refrescamento e migração implicam várias cópias do mesmo documento em intervalos de tempo mais ou menos regulares, pelo que é necessário estabelecer o número de cópias de preservação permitido por lei. Existem ainda outras estratégias, como a migração, em que as sucessivas iterações podem acarretar uma perda de funcionalidade, afectando a

¹⁰⁶ Artigo 221º: “ 1 - As medidas eficazes de carácter tecnológico não devem constituir um obstáculo ao exercício normal pelos beneficiários das utilizações livres previstas nas (...) do n.º 2 do artigo 75.º (...) do Código, no seu interesse directo, devendo os titulares proceder ao depósito legal, junto da Inspecção-Geral das Actividades Culturais (IGAC), dos meios que permitam beneficiar das formas de utilização legalmente permitidas. 2 - Em ordem ao cumprimento do disposto no número anterior, os titulares de direitos devem adoptar adequadas medidas voluntárias, como o estabelecimento e aplicação de acordos entre titulares ou seus representantes e os utilizadores interessados.” 3 - Sempre que se verifique, em razão de omissão de conduta, que uma medida eficaz de carácter tecnológico impede ou restringe o uso ou a fruição de uma utilização livre por parte de um beneficiário que tenha legalmente acesso ao bem protegido, pode o lesado solicitar à IGAC acesso aos meios depositados nos termos do n.º 1. (Portugal, 2008, p. 1930)

integridade do documento. Será, portanto, importante saber se tal acção constitui ou não uma violação do direito moral do autor. Também seria relevante saber se as bibliotecas têm o direito de efectuar uma cópia de preservação dos documentos digitais a que têm acesso através de contrato de licenciamento ou se esse direito tem de ser negociado aquando do acto contratual.

Sem desvalorizar estes aspectos menos claros do *Código de Direito de Autor e Direitos Conexos* português, é de salientar a sua conformidade com as exigências do novo contexto digital.

5.3.1. O direito de autor e a preservação digital nos países da UE

Sobre o impacto do direito de autor na preservação digital destacam-se dois estudos de 2008: *International Study on the Impact Of Copyright law on Digital Preservation* – relatório conjunto de Library of Congress (LC), Joint Information Systems Committee (JISC), Open Access to Knowledge (OAK), Law Project e SURFfoundation – que analisa as leis de direitos de autor da Austrália, Holanda, Reino Unido e Estados Unidos da América e o seu impacto na preservação digital; e *Study on copyright limitations and exceptions for libraries and archives*. Este último estudo, preparado por Kenneth Crews para a WIPO, analisa as legislações sobre os direitos de autor nos 184 países membros e a forma como elas abarcam as excepções ao direito de autor conferidas às bibliotecas e que lhes permitem assegurar o cumprimento dos seus objectivos de investigação e estudo, de preservação e substituição de materiais e de empréstimo interbibliotecário.

As restrições e condições relativas à preservação estabelecidas nas legislações são muitas vezes problemáticas, sobretudo numa época em que as bibliotecas têm de lidar grandes quantidades de informação em formato digital e sentem, por isso, a necessidade de preservar esse património. Atentos a este problema, muitos países formularam novas propostas legislativas para o estabelecimento de práticas que permitissem assegurar a preservação digital. Essas iniciativas são mais evidentes nos países da UE e nos EUA.

As principais conclusões relevantes para a questão da preservação digital do *Study on copyright limitations and exceptions for libraries and archives* assinalam que, dos 149 países cujas legislações foram analisadas, apenas 72 permitem a realização de cópias com fins de preservação; 79 países proíbem a ilusão das medidas de protecção tecnológica; 26 países (21 deles são membros da UE, os outros 5 são a Austrália, a Croácia, os EUA, a Noruega e Singapura) prevêm uma excepção explícita em benefício das bibliotecas; e finalmente 67 países autorizam cópias para substituição. O número de cópias permitido varia consoante a legislação (uma única cópia ou múltiplas); variam também as condições de realização das

cópias (apenas se a consulta a puder danificar a obra ou se esta já se encontra danificada; se é para integrar a colecção permanente da biblioteca; se a obra está indisponível para aquisição/esgotada; se se trata de formatos digitais, etc.).

As leis analisadas abordam a questão dos meios digitais de maneiras diferentes: em alguns países, é permitida a cópia digital; noutros não, ou então só é permitida em alguns casos. Nos países que permitem a cópia digital esta faz-se apenas para efeitos de preservação ou substituição. Neste âmbito, destaca-se a legislação da Estónia, país que, segundo o autor, contém uma das disposições mais agressivas no que diz respeito à cópia digital para fins de preservação:

Its preservation statute includes the explicit right to make a digital collection for preservation purposes. The statute requires that the library search the market for a purchase, but even if the work is available, the library's right to make a digital preservation copy remains. Apparently, the law of Estonia permits libraries to build extensive collections of digital copies, although it is not clear when or whether those copies may be used. (Crews, 2008, p. 53)

Não é objectivo deste trabalho analisar em detalhe as conclusões deste estudo de Crews. Todavia, resultando as legislações sobre os direitos de autor dos países da UE da transposição da *Directiva 2001/29/EC*, será interessante analisar a forma como os 27 países membros, incluindo Portugal, a acolheram no seu direito interino. Para o efeito, foram elaboradas tabelas¹⁰⁷ onde se comparam as principais disposições legais relacionadas especificamente com a preservação digital¹⁰⁸.

Nem todas as legislações se referem claramente a cópias para fins preservação digital, no entanto, considerou-se que as cópias para uso bibliotecário podem e devem integrar as cópias para fins de preservação. Assim, em *1. Cópias para uso bibliotecário* analisa-se, quais as obras cuja cópia é permitida (qualquer obra publicada; obras esgotadas ou não publicadas; obras cujo original pertence à colecção da biblioteca; não especifica¹⁰⁹); que tipo de cópia pode ser feito (reprográfica; digital; não especifica); quantas cópias podem ser feitas (uma só cópia, mais do que uma cópia, não especifica); e qual a finalidade da cópia (cópia de preservação/substituição; não especifica). Em *2. Ilusão de medidas tecnológicas de protecção*, referem-se os actos

¹⁰⁷ Ver Anexo 7

¹⁰⁸ Para uma análise comparativa das limitações e excepções ao Direito de Autor que se encontram presentes nas legislações dos 15 membros mais antigos da EU, ver Dias (2009).

¹⁰⁹ Incluíram-se aqui as legislações que, embora não se referissem especificamente à cópia para fins de preservação, se referem a cópias para uso bibliotecário, cópias que servem os interesses da biblioteca e cópias para uso interno da biblioteca.

proibidos pelas diferentes legislações (acto de ilusão; comercialização de dispositivos; prestação de serviços) e as excepções consagradas às bibliotecas¹¹⁰ (se existem ou não).

Nas legislações dos 26 países da UE contemplados no estudo de Crews, não é clara a distinção entre documentos analógicos e documentos digitais. Mencionam-se as cópias digitais como forma de preservar as colecções analógicas, mas não se clarifica de que modo as colecções digitalizadas e digitais podem ser preservadas legalmente. Não obstante essa ausência de clarificação, deve salientar-se o facto de, em 21 países as respectivas legislações incluírem excepções que permitem às bibliotecas iludirem as medidas tecnológicas de protecção de suma importância para as actividades de preservação digital.

¹¹⁰ A maioria das legislações permitem às bibliotecas iludir as medidas tecnológicas de protecção para beneficiarem das excepções e privilégios concedidos nas legislações. Muitas delas também prevêm que se possa agir judicialmente contra os detentores dos direitos que dificultem tal processo.

Conclusão

A acção de preservação protege a dependência cada vez mais crítica da nossa sociedade em relação à informação e assegura que a informação digital produzida hoje e no passado estará disponível no futuro de forma autêntica. Preservar a informação digital é uma responsabilidade das actuais gerações perante as gerações futuras.

Neste trabalho, pretendeu-se oferecer uma visão geral da preservação da informação digital numa perspectiva orientada para as bibliotecas. Partiu-se da premissa de que as bibliotecas, agora como no passado, assumem um papel cimeiro na preservação, mantendo a informação digital acessível e de forma autêntica. Por isso, é imperioso que os bibliotecários compreendam a especificidade e os desafios colocados por este tipo de preservação. Foi objectivo deste estudo sistematizar os aspectos essenciais da preservação digital, nomeadamente: (i) compreender o valor actual da preservação digital e posicionar as bibliotecas neste novo contexto; (ii) assinalar a particularidade da preservação digital como acção continuada e transversal (diferente, por isso, da preservação tradicional); (iii) sublinhar a especificidade dos documentos digitais e os desafios que colocam no momento da sua preservação; (iv) definir e reflectir sobre os conceitos básicos e imprescindíveis à compreensão cabal da preservação digital e de todas as acções que envolve; (v) enunciar e descrever as principais estratégias, as suas (des)vantagens e a sua aplicabilidade; (vi) apontar os repositórios como as novas instituições de memória; (vii) referir os metadados como informação descritiva e técnica essencial ao processo de preservação; e (viii) reflectir sobre os aspectos legais relacionados com a preservação digital.

O valor e importância que a informação assumiu na sociedade actual justifica a ponderação e execução de acções sérias que permitam preservá-la. Em virtude da importância que a informação adquiriu nas últimas décadas, a informação em formato digital permite agilizar a sua criação/produção, distribuição e disseminação. Este novo formato, pese embora os muitos atractivos que representa, coloca sérios desafios em termos de preservação: os suportes que veiculam a informação digital são mais instáveis; o acesso à informação está dependente de um ambiente tecnológico em constante evolução; a preservação da informação digital exige acções continuadas; as estratégias utilizadas para a preservação da informação digital resultam em acções técnicas e, por isso, complexas; é necessário (re)pensar e (re)definir conceitos como autenticidade, integridade, direitos de autor, depósito legal à luz do contexto digital. Por estas razões, tanto as acções e estratégias como os agentes implicados no processo de preservação digital são distintos das acções e agentes implicados na preservação tradicional. Preservar a informação actualmente é mais do que conservar o legado cultural, literário, artístico e científico

gerado no passado. Preservar a informação digital é mantê-la acessível de forma autêntica através de acções e estratégias contínuas que procuram acima de tudo a preservação do conteúdo em detrimento da preservação do suporte. Assim, e tendo em conta que não se pode preservar a dimensão física do objecto digital a longo prazo, a preservação no contexto digital não passa tanto por conservar o documento original, mas sim por mantê-lo idêntico ao original, preservando as suas propriedades significativas. Neste sentido, conceitos como autenticidade e integridade têm de ser interpretados em conformidade com o novo contexto digital.

A fragilidade e vulnerabilidade de que o património cultural, literário e científico produzido digitalmente se reveste e, simultaneamente, a importância e valor que assume na sociedade actual, não devem demover as tradicionais instituições de memória, particularmente as bibliotecas, da sua missão de preservação. Pelo contrário, devem estimulá-las a executarem essa missão com mais afinco e em conformidade com a especificidade desse novo património, adaptando-se às novas exigências. A preservação em ambiente digital exige medidas que devem ser tomadas desde a génese do documento digital. Assim, as tradicionais instituições de memória deixam de ser os únicos agentes de preservação do património cultural, literário e científico. A preservação do património digital implica todos os agentes envolvidos na criação, distribuição e disseminação da informação digital: produtores de informação; produtores de *hardware* e de *software*; editores/distribuidores; instituições de memória (tradicionais e recentes, como é o caso dos repositórios digitais); governos; consumidores/utilizadores. Por se tratar de uma acção transversal e, por vezes, complexa, a preservação digital obriga à partilha de responsabilidades e à cooperação entre os vários agentes.

A perenidade da informação digital é assegurada através de várias acções e estratégias de preservação que devem ser continuadas. Neste trabalho, mostrou-se que existem acções de preservação que apenas preservam a dimensão física do objecto e que, por isso, devem apenas ser aplicadas a curto/médio prazo: boas práticas de armazenamento e manuseamento; refrescamento; preservação da tecnologia. A continuidade dos documentos digitais exige estratégias mais complexas (a emulação e a migração apresentam-se, porventura, como as mais viáveis) e passíveis de aplicação a médio/longo prazo. Não existe uma estratégia padrão aplicável a todos os tipos de documentos digitais. A escolha da estratégia de preservação adequada é ditada pelo tipo de documento digital e pelas propriedades significativas que se pretende conservar.

A aplicação das estratégias de preservação a documentos digitais pode resultar num processo complexo caso não sejam implementadas medidas que o simplifiquem. Os repositórios digitais surgem como infra-estruturas capazes de regular o processo de ingestão, arquivo, gestão, preservação e acesso às colecções digitais. A normalização dos formatos, a criação de pacotes de informação que incluam os metadados necessários para a descrição, gestão e preservação dos

documentos digitais, a definição de políticas de arquivos e de preservação e a gestão dos direitos de autor e de acesso representam o conjunto de acções executadas por estas novas instituições de memória e que facilitam a preservação do património cultural, literário e científico que se encontra em formato digital. Os repositórios digitais não substituem as bibliotecas: eles estendem significativamente os seus papéis e representam um sério compromisso com a preservação da informação digital.

Outro aspecto crucial abordado neste trabalho foi a questão do depósito legal e dos direitos de autor na era digital. Confrontados com a plasticidade e a vulnerabilidade dos documentos digitais, os titulares dos direitos reforçaram as medidas de protecção, sejam contratuais (recurso a licenças de utilização), sejam tecnológicas (controlo da cópia, encriptação,...). Perante a intensificação da protecção do direito de autor na era digital, as bibliotecas assistiram à diminuição dos privilégios que lhes foram conferidos em ambiente analógico e viram-se impossibilitadas de levar a cabo a sua missão de colecta, divulgação e preservação. É com o objectivo de restaurar os privilégios que permitam às bibliotecas levar a cabo a sua missão que as legislações de vários países têm de ser revistas e actualizadas: só assim podem ser salvaguardados os mesmos trâmites que, em ambiente analógico, permitiram às bibliotecas usufruir das excepções e limitações aos direitos de autor. A actual legislação portuguesa sobre os direitos de autor parece salvaguardar esses privilégios e excepções. No entanto, a lei sobre o depósito legal surge como principal entrave à preservação do património cultural, literário e científico, pois ela não é extensível aos documentos digitais. É prioritário que esta lei seja revista e actualizada em conformidade com o novo contexto. Em complemento deve ser criado projecto nacional de preservação digital, com a constituição, por exemplo, de um repositório digital central¹¹¹ que sirva de depósito para a produção cultural nacional em formato digital e, assim, e que permita preservá-la.

Na sua qualidade de Estado-Membro da União Europeia e de membro de organizações como as Nações Unidas e a WIPO, Portugal deve agir em conformidade com as directrizes e orientações dessas instituições, ajustando a sua legislação e os seus procedimentos às exigências actuais, para que as instituições de memória (em particular, as bibliotecas) assegurem a sua missão de preservação.

¹¹¹ Borbinha (2004) e Pinto (2005) dão conta de um serviço construído na BND – o Repositório de Preservação – que assenta na selecção de um modelo de metadados estruturais e numa solução informática em *open source*, escalável e de custo relativamente reduzido para grandes quantidades de informação. Todas as obras depositadas na BND serão estruturadas segundo o esquema METS (Metadata Encoding and Transmission Standard), incluindo ainda os seus metadados bibliográficos e enviadas para o repositório nessa forma.

Um longo caminho foi percorrido até à data, mas, como lembra Caplan (2008, p. 5), a preservação digital continua a ser uma área jovem, altamente dependente da investigação e da experimentação. Em Portugal, a investigação nesta área e a implementação de projectos de preservação digital parecem não estar ainda tão enraizados e experimentados como nos restantes países da União Europeia. Pese embora a publicação de alguns artigos em que este assunto é abordado do ponto de vista das bibliotecas, essa literatura não oferece dados suficientes para avaliar, por exemplo, de que forma é que a preservação da informação digital está a ser feita nas bibliotecas portuguesas. Esta escassez constituiu uma limitação para a elaboração deste estudo e deve ser colmatada. Neste sentido, seria particularmente interessante saber de que forma as bibliotecas portuguesas estão a preservar as suas colecções digitais. A proliferação de repositórios institucionais, na última década, deixa adivinhar que existe um comprometimento, sobretudo por parte das bibliotecas universitárias, em tornar acessível a produção científica portuguesa. Resta saber se o comprometimento de preservação é a longo prazo. A maior parte do *software* utilizado para a constituição de repositórios digitais oferece medidas de preservação básicas: introdução de metadados; normalização de formatos; actualização e migrações de versões; gestão dos direitos de autor; *backups*; entre outras. No entanto, talvez por se tratarem de processos automatizados, a forma como eles são executados e implementados não tem vindo a ser partilhada e/ou tornada pública. A publicação de documentação sobre esses procedimentos é essencial para a partilha de experiências, de resultados e de políticas em preservação da informação digital. O esforço de cooperação deve começar por aqui.

A integração de repositórios institucionais nas bibliotecas universitárias parece ser uma tendência instituída e isso pode significar que, pelo menos a curto/médio prazo, a informação digital que tutelam (mas apenas um tipo específico de informação – a informação científica) estará preservada. No entanto, o mesmo não parece verificar-se nas bibliotecas públicas. Tal não significa que estas não possuam colecções digitais. Aliás, muitas delas dispuseram de financiamentos para digitalização de algumas das suas colecções. Neste sentido, seria também interessante saber de que forma é que as bibliotecas públicas portuguesas estão a preservar, não só as suas colecções digitalizadas, mas também todos documentos digitais que recebem em quantidades cada vez maiores.

Contactos informais revelaram que, apesar de não possuírem um plano de preservação digital efectivo, as bibliotecas portuguesas executam acções e estratégias de preservação básicas que, no futuro, se podem revelar insuficientes. A pesquisa bibliográfica que foi feita para este trabalho resultou na constatação de um hiato de publicações sobre este assunto em Portugal. Isso pode significar duas coisas: ou as bibliotecas portuguesas não estão a preservar as suas colecções digitais, ou fazem-no sem partilhar experiências nem conhecimentos na área da preservação digital. Tanto uma, como outra resultam em posturas gravosas: por um lado, a

ausência de medidas de preservação que pode resultar na perda irremediável do património digital; por outro lado, a resistência à partilha e à cooperação que são acções fulcrais para o sucesso da preservação digital. A formação de consórcios, a partilha de experiências, de recursos, de custos e a divisão de responsabilidades são actos de cooperação que podem resultar na melhoria de sinergias (e consequentemente, de resultados) com a eliminação de redundâncias e com a anulação de esforços inúteis. Ao cooperarem entre si, ao unirem esforços, as bibliotecas ficam mais fortalecidas no sentido de exercerem importantes influências: junto dos governos, no sentido de estes ajustarem as leis para que as bibliotecas possam prosseguir com a sua missão de curadoras do património cultural, literário e científico; junto dos autores no sentido de seguirem boas práticas aquando da criação de documentos digitais e de, por exemplo, libertarem os documentos digitais de mecanismos tecnológicos de protecção que impeçam a aplicação de estratégias de preservação; junto dos editores/distribuidores, sensibilizando-os no sentido de não sobrepor os seus interesses económicos ao interesse público que a preservação da informação digital representa; junto dos produtores de *software* que para facultem a informação necessária para a implementação das estratégias de preservação digital; etc.

Torna-se, pois, necessária investigação de fundo que permita avaliar a forma como as bibliotecas portuguesas preservam as suas colecções digitais. É necessária também a implementação de medidas de preservação a nível nacional, a criação de guias e manuais de boas práticas, o estabelecimento de consórcios e a partilha de experiências com a publicação de planos e políticas de preservação. Sem estas medidas corre-se o risco de nada estar a ser feito ou de se estarem a duplicar esforços. A preservação do património cultural em formato digital não se coaduna com a inacção e a resistência à cooperação. Este tipo de atitudes pode condenar esse património irremediavelmente.

Anexos

Anexo 1: Boas práticas para o armazenamento e manuseamento de suportes magnéticos e suportes ópticos

SUPORTES MAGNÉTICOS

Manuseamento:

- Reduza o manuseamento dos suportes;
- Evite tocar na superfície de qualquer fita ou disquete magnética. A gordura existente na pele deixa resíduo que pode impedir a leitura e atrair poeiras;
- Não toque na superfície da fita nem nos bordos da fita, a não ser que seja absolutamente necessário. Nesse caso, use luvas de algodão;
- Não use produtos comerciais para limpar fitas e disquetes. Contacte um profissional especializado na limpeza e na reparação de fitas sujas e danificadas;
- Reponha as fitas e as disquetes nas respectivas caixas imediatamente a seguir à sua utilização, para evitar possíveis danos e acumulação de poeira;
- Nunca use clips ou fita adesiva para colocar notas directamente nas cassetes, bobinas ou disquetes;
- Manuseie as fitas em áreas limpas;
- Não deixe as fitas nem as pontas das fitas arrastarem pelo chão;
- Reponha as fitas nas respectivas embalagens quando não estiverem a ser utilizadas;
- Não deixe cair as fitas nem as sujeite a movimentações bruscas;
- Corte as pontas das fitas que estiverem danificadas e enroladas em bobinas;
- Não use fitas adesivas vulgares para segurar o fim da fita ou para fazer emendas. Se necessário, utilize produtos adesivos apropriados para esse efeito.
- Rebobine as fitas periodicamente – não deixe a fita a meio, rebobine sempre a fita até ao fim.

Armazenamento:

- Conserve as fitas magnéticas e as disquetes afastadas de campos magnéticos – não empilhe fitas em cima de equipamento eléctrico;
- Conserve as áreas de armazenamento limpas e sem poeira. A poeira atrai e absorve humidade e acelera a hidrólise, uma causa comum e grave de degradação da fita magnética a longo prazo. A poeira pode também danificar as fitas de forma permanente: o poder abrasivo da poeira, auxiliado pela pressão exercida pelas cabeças do gravador na superfície da fita, risca a camada de óxido e as próprias cabeças do gravador;
- Não deixe bobinas abertas e/ou cassetes expostas a sol;
- Armazene na vertical bobinas abertas e cassetes. As bobinas devem ficar penduradas pela parte central;
- Utilize apenas bobinas e cassetes, caixas, contentores e acessórios de alta qualidade;
- Use cintas de protecção para as fitas de bobinas;
- Não armazene fitas em bolsas ou caixas de cartão de má qualidade, que podem ser ácidas, ou em caixas de vinil que contenham cloro.

Recomendações sobre condições ambientais para o armazenamento:

- As áreas de armazenamento das fitas (bandas) devem ser frescas e secas: 15±3°C e 30-40% HR constituem boas condições de armazenamento. Demasiado calor ou frio danifica os suportes magnéticos;
- A HR superior a 40% acelera a deterioração da camada aglutinadora da fita;
- Evite sujeitar as fitas a mudanças rápidas de temperatura. Se as temperaturas da área de armazenamento e de utilização diferem em mais de 8°C, deixe a fita atingir lentamente a temperatura a que se encontra a sala onde ela vai ser utilizada, mais precisamente, deixe-a repousar 4h para cada 10°C de diferença.

Suportes ópticos

Manuseamento:

- Segure os discos pelo rebordo exterior ou pelo orifício central – não toque na superfície do disco;
- Não dobre o disco;
- Use um marcador de ponta de feltro, de base não solvente, para escrever na etiqueta do disco – não utilize etiquetas adesivas;
- Guarde os discos nas respectivas caixas imediatamente após a sua utilização
- Mantenha os discos nas suas embalagens (ou caixas) para minimizar os efeitos das alterações ambientais;
- Elimine a sujidade, os corpos estranhos, dedadas, manchas e líquidos com um pano de algodão limpo, partindo do centro do disco, em linha recta, em direcção ao rebordo exterior;
- Utilize detergente próprio para CD/DVD, isopropanol ou metanol, para remover sujidades ou corpos estranhos renitentes;
- Verificar a superfície do disco antes de gravar;
- Não deixe os discos nas unidades de leitura;
- Utilize luvas de algodão, sem pêlo, quando manusear material óptico.

Armazenamento:

- Mantenha o disco livre de poeiras ou outro tipo de sujidade;
- Arrume os discos na posição vertical (como os livros) em caixas de plástico especificamente destinadas a CD e DVD;
- Não exponha os discos prolongadamente à luz do sol ou outras fontes de luz ultravioleta;

Recomendações sobre condições ambientais para o armazenamento:

- Conserve os discos em local fresco, seco, escuro e arejado: $20\pm4^{\circ}\text{C}$ e 20-60%HR;
- Em condições de mais calor e mais humidade, produz-se oxidação das camadas metálicas reflectoras, escurecimento e alteração das cores e deterioração das bases poliméricas e do revestimento.
- Uma temperatura de 18°C e uma HR de 40% são considerados valores adequados para armazenamento prolongado. Temperaturas e HR abaixo destes valores são adequados para um armazenamento excepcionalmente longo.

Anexo 2: Factores de sustentabilidade, qualidade e funcionalidade dos formatos digitais

FACTORES DE SUSTENTABILIDADE

A LC distingue sete factores de sustentabilidade: *disclosure* (divulgação); *adoption* (adopção); *transparency* (transparência); *self-documentation* (auto-documentação); *external dependencies* (dependências externas), *impact of patents* (impacto das patentes); e *technical protection mechanisms* (mecanismos de protecção técnica).

Tabela 2: *Factores de sustentabilidade dos formatos.* Fonte: Library of Congress (2007)

Divulgação
Prende-se com o grau de existência e com a acessibilidade às especificações completas e às ferramentas que permitem validar a integridade técnica do objecto digital. A preservação do conteúdo de certos formatos não é praticável sem o conhecimento da forma como a informação foi codificada. Os formatos mais bem documentados são os formatos não-proprietários.
Adopção
Relaciona-se com a percentagem de utilização do formato por parte dos criadores, disseminadores e utilizadores do objecto digital. Um formato mais adoptado não se torna obsoleto tão rapidamente. Além disso, dada a sua larga adopção, as próprias indústrias produzem as ferramentas de migração e emulação, sem que para tal tenha de haver um investimento por parte das organizações de arquivo.
Transparência
Prende-se com o grau em que a representação digital está aberta à análise directa. Aqueles formatos cuja informação oculta é representada de forma simples e directa serão mais fáceis de migrar para outros formatos. A transferência será melhorada se o conteúdo textual empregar codificações de caracteres normalizadas (UNICODE, por exemplo) e se este for armazenado numa forma de leitura natural. A encriptação é incompatível com a transferência e a compressão inibe-a.
Auto-documentação
Deve conter metadados descritivos, técnicos e outros metadados administrativos relacionados com a criação e evolução do ciclo de vida do objecto digital de forma a poder garantir a sua preservação e acesso no futuro.
Dependências externas
Prende-se com o grau com que um dado formato depende de um determinado <i>hardware</i> , sistema operativo ou <i>software</i> para a sua reprodução ou utilização. Um formato dependente acarreta dificuldades em futuros ambientes técnicos que não aqueles de que depende.
Impacto das patentes
Tem a ver com a forma como a existência de patentes vai inibir a capacidade de arquivo por parte das organizações.
Mecanismos de protecção técnica
A implementação de mecanismos como a encriptação impede a preservação e o acesso ao conteúdo do objecto digital no futuro por parte das organizações de arquivo. Para preservar o conteúdo digital e torná-lo acessível, estas organizações devem poder replicar o conteúdo em novos media, migrá-lo e normalizá-lo de forma a enfrentar as mudanças tecnológicas.

FACTORES DE QUALIDADE E FUNCIONALIDADE

Quanto aos factores de qualidade e funcionalidade, eles variam de acordo com o tipo de conteúdo e de acordo com as formas de expressão. Por exemplo, as características de um registo sonoro são diferentes das características duma imagem estática e nem todos os formatos de imagem são apropriados a todos os tipos de imagens. Distinguem-se quatro categorias de conteúdo: imagens estáticas, som, texto e vídeo e apresenta-se, para cada uma delas, os principais factores de qualidade e funcionalidade, segundo a LC:

Tabela 3: *Factores de qualidade e funcionalidade dos formatos.* Fonte: Library of Congress (2007)

Imagens estáticas: <ul style="list-style-type: none">• Claridade (imagem de alta resolução);• Manutenção da cor (gama de cores);• Suporte para efeitos gráficos e tipografia.
Som: <ul style="list-style-type: none">• Fidelidade (som de alta resolução);• Suporte para múltiplos canais (MIDI).
Texto: <ul style="list-style-type: none">• Suporte para a integridade da estrutura do documento e para a navegação;• Suporte para a integridade <i>do layout</i>, fonte e outras características de design;• Suporte para interpretar símbolos matemáticos, fórmulas, diagramas, ...
Vídeo: <ul style="list-style-type: none">• Claridade (imagem de alta resolução);• Fidelidade (som de alta resolução);• Suporte para múltiplos canais de som.

Anexo 3: Formatos adequados para preservação

Os formatos são agrupados em 3 níveis: Alta, Média e Baixa adequação para preservação de: Texto, Imagem, Áudio, Vídeo, Outros.

Tabela 4: *Formatos adequados para preservação.* Fonte: Arquivo da Web Portuguesa (2010)

Texto:

Alta adequação para preservação

- HTML, XHTML ou XML, contendo referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml)
- Textos planos usando codificação de caracteres UTF-8, USASCII ou UTF-16 com Byte Order Mark (.txt)
- PDF/A-1 segundo a norma ISO 19005-1 (.pdf)
- Open Document Text (.odt)

Média adequação para preservação

- HTML, XHTML ou XML, sem referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml)
- Cascading Style Sheets (.css)
- Texto plano com codificação de caracteres em ISO-8859-1 (.txt)
- PDF com fontes embebidas (.pdf)
- Rich Text Format 1.x (.rtf)
- HTML 4.x incluindo declaração de DOCTYPE (.html)
- Open Office Text Document (.sxw)
- Office Open XML (.docx)
- DTD (.dtd)
- SGML (.sgml)

Baixa adequação para preservação

- Microsoft Word (.doc)
- Postscript (.ps)
- PDF cifrado (.pdf)
- WordPerfect (.wpd)
- DVI (.dvi)

Imagem

Alta adequação para preservação

- PNG (.png)
- JPEG2000 usando compressão sem perda de informação (.jp2)
- TIFF sem compressão
- SVG (.svg)

Média adequação para preservação

- JPEG2000 usando compressão com perdas de informação (.jp2)
- GIF (.gif)
- JPEG/JFIF (.jpg)
- TIFF com compressão
- BMP (.bmp)
- Digital Negative (.dng)
- Computer Graphic Metafile e WebCGM (.cgm)

Baixa adequação para preservação

- Macromedia Flash (*.swf)
- PhotoShop (.psd)
- JPEG 2000 Part 2 (.jpf, .jpx)
- MrSID (.sid)
- TIFF em formato Planar (.tiff)
- FlashPix (.fpx)
- RAW
- Encapsulated Postscript (.eps)

Áudio

Alta adequação para preservação

- AIFF com Pulse-code modulation (.aif, .aiff)
- WAV com Pulse-code modulation (.wav, bwf)
- Ogg Vorbis (.ogg, .oga)

Média adequação para preservação

- MP3 (MPEG-1/2, Layer 3) (.mp3)
- Free Lossless Audio Codec (.flac)
- SUN Audio sem compressão (.au)
- Standard MIDI (.mid, midi)
- Advance Audio Coding (.mp4, .m4a, .aac)

Baixa adequação para preservação

- RealNetworks 'Real Audio' (.ra, .rm, .ram)
- Windows Media Audio (.wma)
- WAV com compressão (.wav)
- AIFC com compressão (.aifc)
- NeXT SND(.snd)

Vídeo

Alta adequação para preservação

- QuickTime Movie sem compressão (.mov)
- AVI sem compressão (.avi)
- Motion JPEG 2000 (ISO/IEC 15444-4) (mj2)
- Motion JPEG (.avi, .mov)

Média adequação para preservação

- MPEG-1, MPEG-2 (.mpg, .mpeg)
- MPEG-4 (.mp4)
- Ogg Theora (.ogg, .ogm, .ogv)

Baixa adequação para preservação

- Windows Media Video (.wmv)
- AVI com compressão (.avi)
- QuickTime Movie com compressão (.mov)
- RealNetworks 'Real Video' (.rv, .rm)

Outros formatos

Alta adequação para preservação

- Comma Separated Values (.csv)
- SQL DDL

Média adequação para preservação

- OpenOffice (.sxc/.ods, .sxi/.odp)
- OOXML segundo a norma ISO/IEC DIS 29500 (.xlsx, .pptx)

Baixa adequação para preservação

- *Microsoft Excel (.xls)*
- *Microsoft PowerPoint (.ppt)*
- *Microsoft Access (.mdb)*
- *Microsoft Visio(.vsd)*

Anexo 4: Diagrama esquemático para a selecção das estratégias de preservação digital

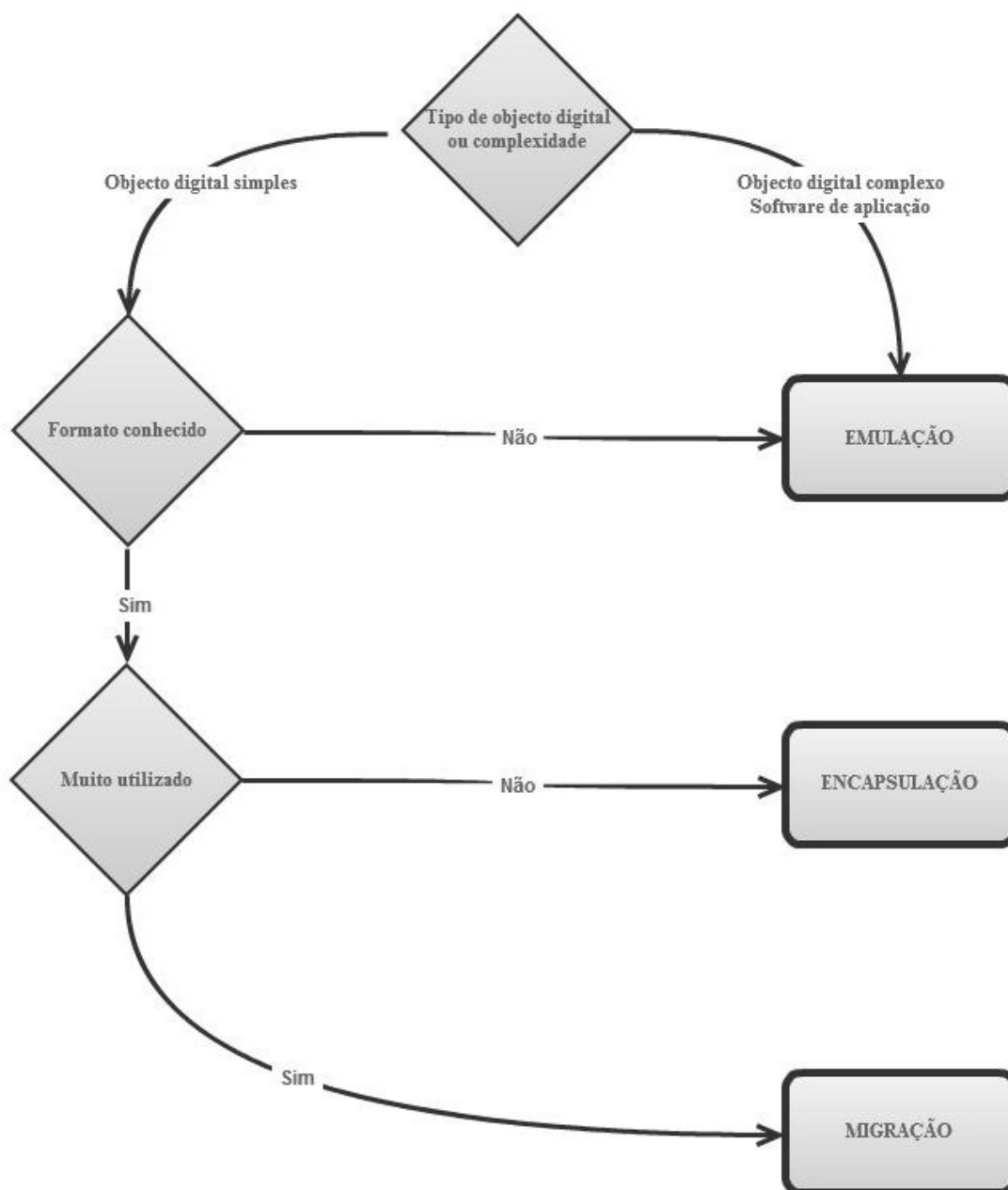


Figura 13: Diagrama esquemático para a selecção das estratégias de preservação digital. Fonte: Lee et al. (2002, p. 103, adaptado)

Anexo 5: Atributos e responsabilidades de um Repositório Digital de Confiança

Atributos de um Repositório Digital de Confiança

- Conformidade com o modelo de referência Open Archival Information System (OAIS)
- Responsabilidade administrativa
- Viabilidade organizacional
- Sustentabilidade financeira
- Adequação tecnológica
- Implementação de sistema de segurança
- Justificação de procedimentos (*Procedural Accountability*)

Responsabilidades de um Repositório Digital Fiável (Trusted Digital Repository)

Responsabilidades de alto nível relativas à organização e manutenção de objectos digitais

- Determinar o objectivo/âmbito das colecções
- Gerir a preservação e o ciclo de vida
- Identificar o conjunto de *stakeholders* detentores dos materiais e *copyright*
- Avaliar custos

Responsabilidades operacionais

- Negociar informação apropriada com os fornecedores de conteúdo
- Adquirir os direitos de autor necessários
- Determinar a comunidade designada
- Assegurar que a informação a ser preservada é compreendida de forma independente por parte da comunidade designada
- Seguir políticas e procedimentos documentados
- Tornar a informação preservada disponível à Comunidade Designada
- Adoptar boas práticas na criação de recursos digitais

Fonte: RLG/OCLC, (2002)

Anexo 6: Modelo de referência OAIS – Pacotes de Informação

Um Pacote de Informação (PI) contém dois tipos de informação: Informação (*content information*) e Informação Descritiva de Preservação (IDP) (*preservation description information*). O primeiro tipo, refere-se à informação foco de preservação, ou seja, à informação em si, ao conteúdo informacional do contido no objecto digital. O segundo, aplica-se ao primeiro e é necessário para descrever essa informação, para assegurar que esta está claramente identificada e para compreender o ambiente em que ela foi gerada. A IDP inclui cinco tipos de informação de preservação¹¹²: (i) referência (fornece um ou mais identificadores, ou sistemas de identificadores, que identificam inequivocamente a informação); (ii) contexto (descreve de que forma a informação se relaciona com outra informação externa ao PI); (iii) proveniência (documenta a história da informação, indicando a sua origem ou fonte, as alterações que sofreu desde que foi criada, quem deteve a sua custódia desde que foi produzida, assegurando, assim, a sua autenticidade); (iv) fixidez (fornece o mecanismo que protege a informação de alterações não documentadas, assegurando a sua integridade); (v) direitos de acesso (fornece os termos de acesso à informação e restrições, que incluem as condições de acesso e distribuição expressas no Acordo de Submissão, relacionadas com a preservação e utilização da informação pelo consumidor final). Informação e IDP aparecem encapsulados num único PI:

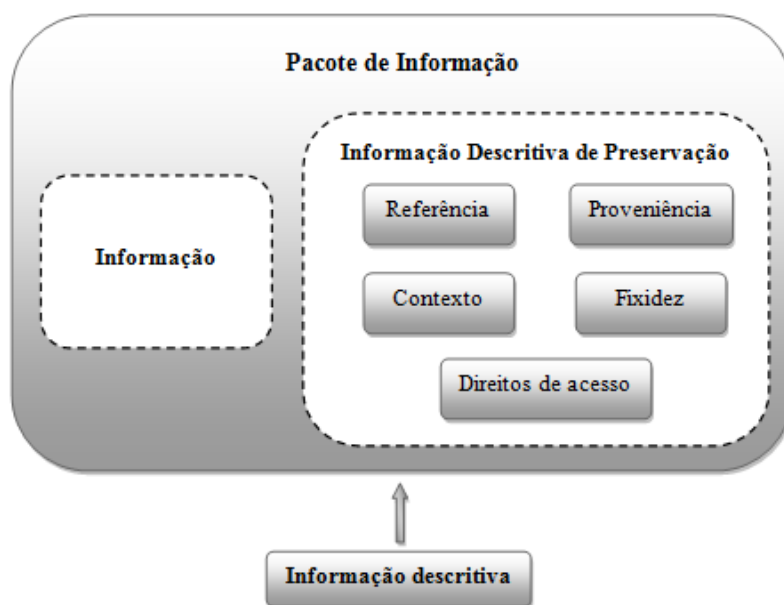


Figura 14: Conteúdo de um Pacote de Informação, segundo o modelo de referência OAIS.

¹¹² Na versão de 2002, são apresentados apenas os quatro primeiros tipos de PDI, sendo os direitos de acesso um novo tipo de PDI, introduzido na revisão de 2009, ainda não aprovada.

A informação descritiva corresponde àquela informação utilizada para descobrir que pacote contém a informação de interesse.

Dependendo do local dentro do ambiente OAIS onde estão a ser utilizados, distinguem-se três tipos de PI: Pacote de Informação para Submissão (PIS), Pacote de Informação para Arquivo (PIA), Pacote de informação para Disseminação (PID).

- PIS: versão do PI enviada pelo Produtor para o OAIS. A forma como a informação é submetida normalmente é negociada entre o Produtor e o OAIS através de um Acordo de Submissão. O conceito de PIS enfatiza o facto de a informação poder ser preservada num formato diferente daquele em que foi submetido.
- PIA: v do PI preparada para ser armazenada e preservada no OAIS. Consiste na informação que é o foco da preservação bem como os metadados necessários para suportar os serviços de preservação e acesso. O PIA estará sempre se acordo com as normas internas do OAIS.
- PID: versão do PI entregue ao consumidor como resultado de uma pesquisa, já em contexto de acesso. O PID enfatiza o facto de este tipo de PI poder ser diferente, quer no formato, quer no conteúdo, do PIA.

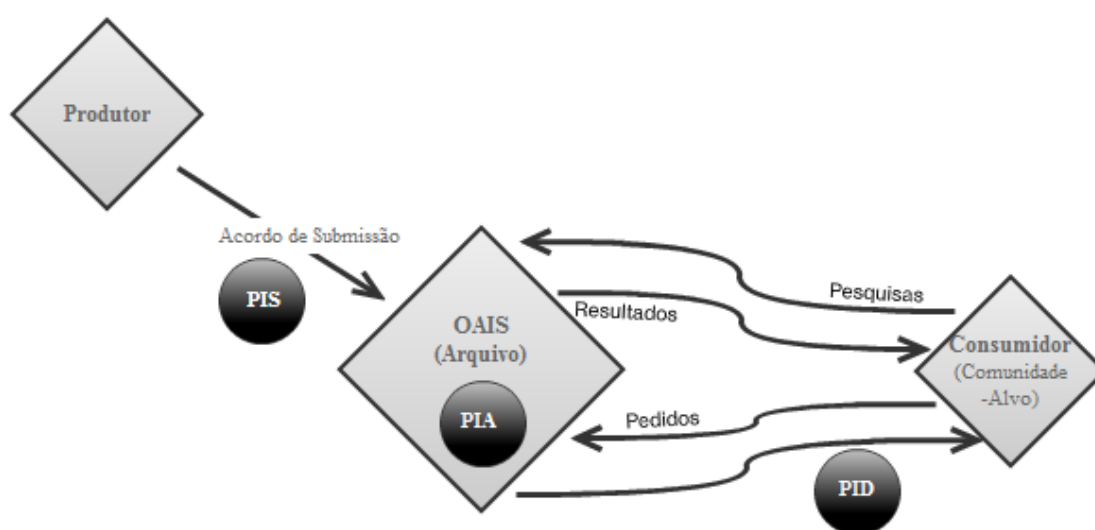


Figura 15: Fluxo dos Pacotes de Informação.

PIS, PIA e PID representam objectos de informação depositados, geridos e difundidos por um arquivo OAIS. O PIS pode não ser igual ao PID, pois não só o processo de arquivo pode ter dado origem a um novo PI, como também o acordo de submissão pode ter estipulado que nem a informação a ser preservada podia ser tornada acessível. Num repositório que segue o

modelo de referência OAIS, os esforços de preservação vão incidir sobre o PIA. O PIA, como qualquer outro pacote de informação, deve incluir, para além da Informação em si, a IDP. A IDP inclui a informação que assegura a confiança na informação a que está associada, permite o seu acesso e contextualiza-a. A IDP descreve ainda os estados passados e presentes da informação depositada, assegurando que ela está identificada de forma única e inequívoca e que não foi alterada.

Anexo 7: Análise do direito de autor e a preservação digital nos países da UE

Tabela 5: Análise da legislação dos países da UE: cópias para uso bibliotecário (natureza das obras)

DISPOSIÇÕES					
1. Cópias para uso bibliotecário					
1.1. Que obras se podem copiar					
	Qualquer obra publicada	Obras esgotadas ou não publicadas	Obras cujo original pertence à colecção da biblioteca	Não específica	Outras disposições
Alemanha				X	
Áustria	X	X (uma só cópia)			
Bélgica	X				
Bulgária	X				
Chipre	X				
Dinamarca	X				
Eslováquia	X		X		
Eslovénia	X				
Espanha	X				
Estónia			X		
Finlândia			X		
França	X				
Grécia			X		
Hungria				X	
Irlanda			X		
Itália			X		
Letónia	X				
Lituânia			X		
Luxemburgo	X				
Malta	X				
Países Baixos			X		
Polónia	X				
Portugal	X				
Reino Unido			X		
Rep. Checa	X				
Roménia		Legislação não contemplada no estudo.			
Suécia	X				

Tabela 6: *Análise da legislação dos países da UE: cópias para uso bibliotecário (tipo e número de cópias)*

DISPOSIÇÕES						
1. Cópias para uso bibliotecário						
1.2. Tipo da cópia			1.3. Número de cópias			
	Reprográfica	Digital	Não específica	Uma só cópia	Mais do que uma cópia	Não específica
Alemanha			X			X
Áustria	X	Apenas se o original pertencer à biblioteca e com fins não comerciais		X		
Bélgica			X		X	
Bulgária			X			X
Chipre			X			X
Dinamarca			X			X
Eslováquia			X			X
Eslovénia			X		X	
Espanha	X	X				X
Estónia		X				X
Finlândia			X			X
França			X			X
Grécia			X	X		
Hungria	X	X		X		
Irlanda			X			X
Itália	X					X
Letónia	X	X		X		
Lituânia	X					X
Luxemburgo			X			X
Malta	X	X			X	
Países Baixos			X			X
Polónia			X			X
Portugal			X			X
Reino Unido	X	X				X
Rep. Checa	X	X				X
Roménia	Legislação não contemplada no estudo.					
Suécia			X			X

Tabela 7: *Análise da legislação dos países da UE: cópias para uso bibliotecário (finalidade da cópia)*

DISPOSIÇÕES				
1. Cópias para uso bibliotecário				
1.4. Finalidade da cópia				
	Cópia de preservação/substituição	Não específica	Condições	Outras disposições
Alemanha		X		
Áustria		X		
Bélgica	X			
Bulgária	X			
Chipre	X			
Dinamarca	X			
Eslováquia	X			
Eslovénia	X			
Espanha	X			
Estónia	X			
Finlândia	X			
França	X			
Grécia		X		
Hungria		X		
Irlanda	X			
Itália		X		
Letónia	X			
Lituânia	X			Exclui obras disponibilizadas ao público através de redes informáticas
Luxemburgo	X			
Malta		X		
Países Baixos	X			
Polónia	X			
Portugal		X		
Reino Unido	X			
Rep. Checa	X			
Roménia		Legislação não contemplada no estudo.		
Suécia	X			

Tabela 8: *Análise da legislação dos países da UE: ilusão das medidas tecnológicas de protecção.*

DISPOSIÇÕES						
2. Ilusão de medidas tecnológicas de protecção						
	2.1. Actos proibidos			2.2. Prevê excepções em benefício das bibliotecas?		
	Acto de ilusão	Comercialização de dispositivos	Prestação de serviços	Sim	Não	Outras disposições
Alemanha	X	X	X	X		
Áustria	X	X	X	X		
Bélgica	X	X	X	X		
Bulgária		X	X			Só com o consentimento do titular dos direitos
Chipre		X	X	X		
Dinamarca	X	X	X	X		
Eslováquia	X	X	X		X	
Eslovénia	X	X		X		
Espanha	X	X	X	X		
Estónia		X		X		
Finlândia	X	X	X	X		
França	X	X	X	X		
Grécia	X	X	X	X		
Hungria	X	X	X	X		
Irlanda		X	X	X		
Itália	X	X	X	X		
Letónia	X	X	X	X		
Lituânia	X	X	X	X		
Luxemburgo	X	X	X	X		
Malta	X	X	X	X		
Países Baixos	X	X	X	X		
Polónia	X	X			X	
Portugal	X	X	X	X		
Reino Unido	X	X	X	X		
Rep. Checa	X	X	X		X	
Roménia		Legislação não contemplada no estudo.				
Suécia	X	X	X	X		

Referências bibliográficas

- ABID, Abdelaziz (2007) – Safeguarding our digital heritage : a new preservation paradigm [Em linha]. In LUSENET, Yola de ; WINTERMANS, Vicent (ed.) – *Preserving the digital heritage : principles and policies*. Hage : UNESCO, 2007. [Acedido em 2010-03-30]. Disponível na WWW: <http://www.knaw.nl/ecpa/publ/pdf/2735.pdf> . ISBN 978-90-6984-523-4.
- ALA (American Library Association) (2007) – *Definitions of digital preservation* [Em linha]. Chicago: ALA. [Acedido em 2010-03-30]. Disponível na WWW: <http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf>
- APBAD (Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas) (2002) – *Tomada de posição e Declaração do Conselho Directivo Nacional de BAD sobre o projecto de diploma destinado a regulamentar a aplicação da Lei do Depósito Legal (21/12/2002)* [Em linha]. [S.l. : s.n.]. [Acedido em 2010-05-05] Disponível na WWW: <http://www.apbad.pt/downloads/tpDeposito.pdf>
- ANGEL-ARELLANO, Miguel (2004) – Preservação de documentos digitais [Em linha]. *Ciência da Informação*. Vol. 33, nº 2 (2004), p. 15-27. [Acedido em 2010-08-30]. Disponível na WWW: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000200002&lng=en&nrm=iso. ISSN 01001965.
- ARMS, Caroline ; FLEISCHHAUER, Carl (2005) – *Digital formats for sustainability, functionality, and quality* [Em linha]. Washington : [s.n.], 2005. [Acedido em 2010-03-30] Disponível na WWW: http://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf
- BARBEDO, Francisco ; RAMALHO, José Carlos ; Ferreira, Miguel (2007) – RODA : Repositório de Objectos Digitais Autênticos. In CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 9, Ponta Delgada, Portugal, 2007 – *Bibliotecas e arquivos : informação para a cidadania, o desenvolvimento e a inovação : actas* [Em linha]. [S.l.] : Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas, 2007. [Acedido em 2011-01-06] Disponível na WWW: <http://repositorium.sdum.uminho.pt/handle/1822/6330>
- BEAGRIE, Neil ; JONES, Maggie (2008) – *Preservation Management of Digital Materials: The Handbook*. [Em linha]. [S.l.] : DPC. [Acedido em 2009-10-28]. Disponível na WWW: <http://www.dpconline.org/advice/digital-preservation-handbook.html>
- BESEK, June M. (2008) – Digital preservation and copyright: an international study [Em linha]. *The International Journal of Digital Curation*. Vol. 3, nº 2 (2008), p. 104-111. [Acedido em 2010-01-03]. Disponível na WWW: <http://www.ijdc.net/index.php/ijdc/article/viewFile/90/61>. ISSN 1746-8256.
- BLUE RÍBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS (2010) – *Sustainable economics for a digital planet: ensuring long term access to digital information* [Em linha]. [S.l. : s.n.]. [Acedido em 2010-03-29]. Disponível na WWW: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

- BÖRGHOFF, U.M. [et al.] (2005) – *Long-term preservation of digital documents: principles and practices*. New York : Springer. ISBN 978-3-540-33639-7.
- BORBINHA, José (2004) – Depósito e preservação na Biblioteca Nacional Digital. In CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8, Estoril, 2004 – *Nas encruzilhadas da informação e da cultura: (re)inventar a profissão: actas* [Em linha]. Lisboa: B.A.D., 2004. [Acedido em 2010-07-23] Disponível na WWW: <http://badinfo.apbad.pt/congresso8/comm8.pdf>.
- BRADLEY, Kevin (2007) – Defining digital sustainability. *Library Trends*. ISSN 1559-0682. Vol. 56, nº 1 (Summer 2007), p. 147-163.
- BYERS, Fred R. (2009) – *Discos ópticos : cuidados e manuseamento : guia para bibliotecários e arquivistas*. Lisboa: BN. ISBN 978-972-565-438-2.
- CANDÁS ROMERO, Jorge (2006) – El papel de los metadatos em la preservación digital. *El Profesional de la Información*. ISSN 1386-6710. Vol. 15, nº 2 (Marzo/Abril 2006), p. 126-136.
- CAPLAN, Priscilla (2008) – The preservation of digital materials. *Library Technology Reports*. ISSN 0024-2586. Vol. 44, nº 2 (February/March 2008).
- CAPLAN, Priscilla (2009) – *Understanding PREMIS* [Em linha]. [S.l.] : LC. [Acedido em 2010-12-16]. Disponível na WWW: <http://www.loc.gov/standards/premis/understanding-premis.pdf>
- CCSDS (Consultative Committee for Space Data Systems) (2002) – *Reference Model for an Open Archival Information System (OAIS): Blue Book, Issue 1* [Em linha]. Washington, DC (US): CCSDS. [Acedido em 2010-03-31]. Disponível na WWW: <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- CCSDS (Consultative Committee for Space Data Systems) (2009) – *Reference Model for an Open Archival Information System (OAIS): Blue Book, Issue 1* [Em linha]. Washington, DC (US): CCSDS. [Acedido em 2010-03-31]. Disponível na WWW: <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- CONWAY, Paul (1996) – *Preservation in the Digital World*. [Em linha]. [S.l. : s.n.]. [Acedido em 2010-07-28]. Disponível na WWW: <http://www.clir.org/pubs/reports/conway2/>
- CREWS, Kenneth (2008) – *Study on copyright limitations and exceptions for libraries and archives* [Em linha]. Geneve : WIPO. Acedido em 2010-04-19]. Disponível na WWW: http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=109192
- CROFT, Jean Ann (2003) – The preservation evolution: a review of the preservation literature: 1999-2001. *Library Resources & Technical Services*. ISSN 0024-2527. Vol. 47, nº 2 (2003), p. 59-70.
- CUNHA, Jacqueline de Araújo ; Lima, Marcos Galindo (2007) – Preservação digital : o estado da arte [Em linha]. In ENANCIB, 8, Bahia, 28 a 31 de Outubro – *Encontro Nacional de Pesquisa em Ciência da Informação: organização e representação do conhecimento*. Bahia : ENANCIB. [Acedido em 2010-03-31]. Disponível na WWW: <http://www.enancib.ppgci.ufba.br/artigos/GT2--043.pdf>

- DIAS, Maria do Carmo Ferreira ; FÉRNANDEZ MOLINA, Juan Carlos ; BORGES, Maria Manuel (2009) – Os privilégios das bibliotecas nas leis de direito de autor nos apíses da União Europeia : análise comparativa. In BORGES, Maria Manuel ; SANZ CASADO, Elias (coord.) – *A ciência da informação criadora de conhecimento: Actas do IV Encontro Ibérico EDIBCIC 2009* [Em linha]. Universidade de Coimbra. Vol. 2 ISBN. Disponível na WWW: http://www.eventos-iuc.com/ocs/public/conferences/1/schedConfs/1/actas_EDIBCIC2009_2.pdf. ISBN 978-989-26-0014-7.
- DURANTI, Luciana (2001) – Concepts, principles, and methods for the management of electronic records. *The Information Society*. ISSN 0197-2243. Vol.17 (2001), p. 271-279.
- FARIA, Maria Isabel ; PERICÃO, Maria da Graça (2008) – *Dicionário do livro : da escrita ao livro electrónico*. Coimbra : Almedina.
- FERNÁNDEZ-MOLINA, J. C.; GUIMARÃES, J.A.C. (2007) – Las nuevas leyes de derecho de autor: ¿adecuadas para la preservación digital? [Em linha]. *Information Research*. ISSN 1368-1613. Vol. 12, nº. 4. [Acedido em 2009-10-28]. Disponível na WWW: <http://informationr.net/ir/12-4/paper322.html>
- FERREIRA, Miguel (2006) – *Introdução à Preservação Digital : Conceitos, estratégias e actuais consensos* [Em linha]. Guimarães : Escola de Engenharia da Universidade do Minho, [Acedido em 2010-03-06]. Disponível na WWW: <http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. ISBN 9728692307.
- FERREIRA, Miguel (2009) – *Preservação de longa duração de informação digital no contexto de um arquivo histórico*. [Em linha]. Guimarães : Escola de Engenharia da Universidade do Minho, [Acedido em 2010-03-06]. Disponível na WWW: <http://repositorium.sdum.uminho.pt/handle/1822/9563>.
- FCCN (Fundação para a Computação Científica Nacional) (2011) – *Arquivo da Web Portuguesa* [Em linha]. [S.l. : s.n.]. [Acedido em 2010-10-28]. Disponível na WWW: <http://arquivo-web.fccn.pt/>
- GANTZ, John F. (2008) – *O diversificado e crescente universo digital : uma previsão atualizada do crescimento mundial das informações até 2011* [Em linha]. [S.l.] : IDC ; EMC. [Acedido em 2010-04-08]. Disponível na WWW: <http://portugal.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>
- GLADNEY, Henry M. (2007) – *Preserving digital information*. New York : Springer. ISBN 978-3-540-37886-0.
- HERMINGER, Alan R.; ROBERTSON, Steven (1998) – Digital Rosetta Stone: A conceptual model for maintaining long-term access to digital documents [Em linha]. IN DELOS Workshop Preservation of Digital Information, 6, Tomar, June 1998 – *Proceedings of the Sixth DELOS Workshop on Preservation of Digital Information*. [S.l.] : ERCIM. [Acedido em 2010-01-06]. Disponível na WWW: <http://eprints.kfupm.edu.sa/35017/1/35017.pdf>
- HOORENS, Stijn [et al.] (2007) – *Addressing the uncertain future of preserving the past. Towards a robust strategy for digital archiving and preservation* [Em linha]. Cambridge :

- RAND. [Acedido em 2010-03-024]. Disponível na WWW: http://www.rand.org/pubs/technical_reports/TR510/
- HOCKX-YU, Helen (2006) – Digital preservation in the context of institutional repositories [Em linha]. *Program: electronic library and information systems*. Vol. 40, nº3 (2006), p. 232 – 243. [Acedido em 2009-12-21]. Disponível na WWW: <http://eprints.rclis.org/7351/> ISSN 0033-0337.
- HORRELL, Jeffrey (2008) – Converting and preserving the scholarly record: an overview. *Library Resources & Technical Services*. ISSN 0024-2527 . Vol. 52, nº1 (2008), p. 27-32.
- IFLA (2004a) – *Directrizes da IFLA para a conservação e o manuseamento de documentos de biblioteca*. Lisboa: BN. ISBN 972-565-306-8.
- IFLA (2004b) – *Limitations and Exceptions to Copyright and Neighbouring Rights in the Digital Environment: An International Library Perspective* [Em linha]. Buenos Aires: IFLA. [Acedido em 2010-02-03]. Disponível na WWW: <http://archive.ifla.org/III/clm/p1/ilp.pdf>
- IFLA ; WIPO (2009) – *Statement of Principles on Copyright Exceptions and Limitations for Libraries and Archives* [Em linha]. Geneva: IFLA. [Acedido em 2010-02-03]. Disponível na WWW: <http://www.ifla.org/files/clm/statements/statement-of-principles-sccr20.pdf>
- JORDAN, Sophia K. (2000) – A review of the preservation literature: 1993-1998. *Library Resources & Technical Services*. ISSN 0024-2527. Vol. 44, nº 1 (2000), p. 4-21.
- KEEFER, Alice ; GALLART, Núria (2007) – *La preservación de recursos digitales : el reto para las bibliotecas del siglo XXI*. Barcelona: UOC. ISBN 978-84-9788-567-6.
- KENNEY, Anne R.et al. (2003) – *Digital preservation management: implementing short-term strategies for long term problems* [Em linha]. [S.l. : Cornell University]. [Acedido em 2009-03-29]. Disponível na WWW: <http://www.icpsr.umich.edu/dpm/>
- KUNY, Terry (1997) - A Digital Dark Ages? Challenges in the Preservation of Electronic Information. In IFLA Council and General Conference, 63, 1997 – *IFLA Conference Proceedings* [Em linha]. S.l. : IFLA, 1997. [Acedido em 2010-03-06]. Disponível na WWW: <http://archive.ifla.org/IV/ifla63/63kuny1.pdf>
- LARIVIÈRE, Jules (2000) – *Guidelines for legal deposit legislation* [Em linha]. Paris: UNESCO. [Acedido em 2010-02-08]. Disponível na WWW: http://portal.unesco.org/ci/en/ev.php-URL_ID=24108&URL_DO=DO_TOPIC&URL_SECTION=201.html
- LAVOIE, Brian F. (2004) – *The Open Archival Information System Reference Model : Introductory Guide* [Em linha]. Dublin, USA: OCLC. [Acedido em 2010-09-06]. Disponível na WWW: http://www.dpconline.org/docs/lavoie_OAIS.pdf
- LAVOIE, Brian ; DEMPSEY, Lorcan (2004) – Thirteen ways of looking at... Digital Preservation [Em linha]. *D-Lib Magazine*. Vol. 10, nº 7/8 (July/August 2004). [Acedido em 2010-03-05]. Disponível na WWW: <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>. ISSN 1082-9873.

- LAVOIE, Brian ; GARTNER, Richard (2005) – *Preservation metadata : technology watch report* [Em linha]. [S.l.] : OCLC/DPC. [Acedido em 2010-06-15]. Disponível na WWW: <http://www.dpconline.org/docs/reports/dpctw05-01.pdf>
- LAZINGER, Susan S. (2001) – *Digital preservation and metadata : history, theory, practice*. Colorado : Libraries Unlimited. ISBN 1-56308-77-4.
- LEE, Kyong-Ho (2002) – The state of the art and practice in digital preservation [Em linha]. *Journal of Research of the National Institute of Standards and Technology*. [Acedido em 2010-03-05]. Disponível na WWW: <http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf> ISSN 1044-93-106.
- LIBRARY CONGRESS (2010) *Digital Preservation* [Em linha]. [S.l. : s.n.] [Acedido em 2010-02-17]. Disponível na WWW: <http://www.digitalpreservation.gov/>.
- LUSENET, Yola de (2002) – Digital heritage for the future. *Cadernos BAD*. ISSN 0007-9421. 2 (2002)
- MARCUM, Deanna B. (1996) – The preservation of digital information. *The Journal of Academic Librarianship*. ISSN 0099-1333. Vol. 22, nº 6 (November 1996), p. 451-454.
- MARCUM, Deanna B. (2003) – Research questions for the digital era library. *Library trends*. ISSN 0024-2594. Vol. 51, nº 4 (Spring 2003), p. 636-651.
- MARCUM, Deanna ; FRIEDLANDER, Amy (2003) - Keepers of the crumbling culture: what digital preservation can learn from library history [Em linha]. *D-Lib Magazine*. Vol. 9, nº 5 (May 2003) [Acedido em 2010-02-17]. Disponível na WWW: <http://www.dlib.org/dlib/may03/friedlander/05friedlander.html>. ISSN 1082-9873.
- MÁRDERO ARELLANO, Miguel Ángel (2008) - *Critérios para a preservação digital da informação científica* [Em linha]. Brasília : Universidade de Brasília. [Acedido em 2010-03-29]. Disponível na WWW: <http://repositorio.bce.unb.br/handle/10482/1518>
- MÁRDERO ARELLANO, Miguel Ángel ; Leite, Fernando César Lima (2009) - Acesso aberto à informação científica e o problema da preservação digital [Em linha]. *Biblos*. Nº 35 (Mar.-Jun., 2009), p. 1-11. [Acedido em 2010-03-07]. Disponível na WWW: <http://www.scielo.org.pe/pdf/biblios/n35/a01n35.pdf> . ISSN 1562-4730.
- MASON, Ingrid (2007) – Virtual preservation: How has digital culture influenced our ideas about permanence? Changing practice in a national legal deposit library *Library trends*. ISSN 0024-2594. Vol. 56, nº 1 (Spring 2007), p. 198-215.
- MEMORY OF WORLD PROGRAMME (2005) – *Preserving our documentary heritage* [Em linha]. [S.l.] : Unesco. [Acedido em 2010-03-24]. Disponível na WWW: http://portal.unesco.org/ci/en/ev.php-URL_ID=19440&URL_DO=DO_TOPIC&URL_SECTION=201.html
- MÉNDEZ RODRÍGUEZ, Eva Maria (2002) – *Metadatos y recuperación de información : estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón : Trea. ISBN 84-9704-055-4.

- MOGHADDAM, Golnessa Galsyani (2007) – Archiving Challenges of Scholarly Electronic Journals: How Do Publishers Manage Them?. *Serials Review*. ISSN 0098-7913. Vol. 33, nº 2 (2007), p. 81-90.
- MONTFORT, Nick ; WARDROP-FRUIIN, Noah (2004) – *Acid-free bits : recommendations for long-lasting electronic literature* [Em linha]. [S.l.] : Electronic Literature Organization, 2004 [Acedido em 2010-07-15]. Disponível na WWW: <http://eliterature.org/pad/afb.html>
- MUIR, Adrienne (2004) – Digital preservation: awareness, responsibility and rights issues. *Journal of Information Science*. ISSN 0165-5515. Vol. 30, nº 1 (2004) , p. 73-92.
- NISO (National Information Standards Organization) (2004) – *Understanding metadata* [Em linha]. Bethesda : NISO Press. [Acedido em 2010-04-06]. Disponível na WWW: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>. ISBN 1-880124-62-9.
- OWEN, John Mackenzie (2007) – Preserving the digital heritage : roles and responsibilities for heritage repositories [Em linha]. In LUSENET, Yola de ; WINTERMANS, Vicent (ed.) – *Preserving the digital heritage : principles and policies*. Hage : UNESCO, 2007. [Acedido em 2010-03-30]. Disponível na WWW: <http://www.knaw.nl/ecpa/publ/pdf/2735.pdf> . ISBN 978-90-6984-523-4
- PEARSON, David; WEBB, Colin (2008) - Defining File Format Obsolescence: A Risky Journey [Em linha]. *The International Journal of Digital Curation*. Vol. 3, nº 1 (2008), p. 89-106. [Acedido em 2010-01-08]. Disponível na WWW: <http://www.ijdc.net/index.php/ijdc/article/viewFile/76/44>. ISSN 1746-8256.
- PINTO, Maria Manuela Gomes de Azevedo (2005) – Do «efêmero» ao «sistema de informação» a preservação na era digital [Em linha]. *Páginas a&b*. Nº 15 (2005), p.63-178. [Acedido em 2010-04-05]. Disponível na WWW: <http://ler.letras.up.pt/uploads/ficheiros/3083.pdf>. ISSN 0873-5670.
- PINTO, Maria Manuela Gomes de Azevedo (2007) – Gestão da informação e preservação digital : uma perspectiva portuguesa de uma mudança de paradigma. In CONGRESS ISKO-SPAIN = CONGRESO ISKO-SPAIN, 9, Valencia, 2009 – *New perspectives for the organization and dissemination of knowlwdge = Nuevas perspectivas para la difusión y organización del conocimiento : actas*. [Em linha], Congress Isko-Spain = Congreso Isko-España, 2009. València : Universitat Politècnica de València. [Acedido em 2010-04-05]. Disponível na WWW: <http://repositorio-aberto.up.pt/bitstream/10216/25380/2/manuelapintogestao000100395.pdf>
- PINTO, Maria Manuela Gomes de Azevedo (2007) – *Preservmap : um roteiro da preservação na Era Digital*. Porto : [s.n.].
- PORTUGAL. Assembleia da República (2008) – Lei nº 16/2008 de 1 de Abril [Em linha]. In *Diário da República*. 1ª Série. Nº 64 (1 de Abril de 2008). [Acedido em 2010-01-27]. Disponível na WWW: <http://www.parlamento.pt/ActividadeParlamentar/Paginas/DetalheDiplomaAprovado.aspx?BID=14990>

- PORTUGAL. Conselho de Ministros (2002?) – [Texto do Projecto de Decreto-Lei relativo ao depósito legal] [Em linha]. [Acedido em 2010-05-05]. Disponível na WWW: <http://www.apbad.pt/downloads/projectoDLdepositolegal.pdf>
- PORTUGAL. Ministério da Cultura (2000) – Decreto-lei nº 122/2000 de 4 de Julho [Em linha]. In *Diário da República*. 1ª Série-A. Nº 152 (4 de Julho de 2000). [Acedido em 2010-05-05]. Disponível na WWW: <http://www.unic.pt/images/stories/publicacoes/DL%20122%202000.pdf>
- PORTUGAL. Ministério da Cultura e Coordenação Científica (1982) – Decreto-lei nº 74/82 de 3 de Março [Em linha]. In *Diário da República*. 1ª Série. Nº 51 (3 de Março de 1982). [Acedido em 2010-01-27]. Disponível na WWW: http://www.apimprensa.pt/cache/get_bin.aspx?404;http://www.apimprensa.pt/cache/bin/XPQ4wXwXX182pPjjs8mG4nZKU.pdf
- PORTUGAL. Ministério da Educação e Cultura (1986) – Decreto-lei nº 362/86 de 28 de Outubro [Em linha]. In *Diário da República*. 1ª Série. Nº 249 (28 de Outubro de 1986). [Acedido em 2010-05-05]. Disponível na WWW: http://www.apbad.pt/downloads/dl362_86.pdf
- RLG ; OCLC (2002) – *Trusted Digital Repositories: Attributes and Responsibilities : an RLG-OCLC Report* [Em linha]. Mountain View, CA: RLG. [Acedido em 2010-11-23]. Disponível na WWW: <http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf>
- ROSSI, Christian (2008) - *From distribution to preservation of digital documents* [Em linha]. [S.l. : s.n.] [Acedido em 2010-01-20]. Disponível na WWW: <http://eprints.rclis.org/16334/3/preservation.html>
- SARAMAGO, Maria de Lurdes (2004) – Metadados para a preservação digital e a aplicação do modelo OAIS [Em linha]. In CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 8, Estoril, 2004 – *Nas encruzilhadas da informação e da cultura: (re)inventar a profissão: actas* [Em linha]. Lisboa: B.A.D., 2004 [Acedido em 2010-04-05]. Disponível na WWW: <http://badinfo.apbad.pt/congresso8/comm2.pdf>
- SARAMAGO, Maria de Lurdes (2002) – Preservação digital a longo prazo : boas práticas e estratégias. *Cadernos BAD*. Lisboa: BAD. ISSN 0007-9421. Vol. 2 (2002), p. 54-68
- SENSO, José A. ; ROSA PIÑERO, Antonio de la (2003) – El concepto de metadato : algo más que descripción de recursos electrónicos [Em linha]. *Ciência da Informação*. Vol. 32, nº 2 (Maio/Agosto 2003), p. 95-106. [Acedido em 2010-04-06]. Disponível na WWW: <http://www.scielo.br/pdf/ci/v32n2/17038.pdf>. ISSN 1518-8353.
- STRODL, Stephan [et al.] (2007) – How to choose a digital preservation strategy: evaluating a preservation planning procedure. In LARSON, Ray R. ed. lit. (2007) – *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*. New York: ACM. ISBN:978-1-59593-644-8. P. 29-38.
- SMITH, Abby (2007) – Valuing preservation. *Library Trends*. ISSN 1559-0682. Vol. 56, nº 1 (Summer 2007), p. 4-25.

- SPENCE, Jacqueline (2006) – Preserving the cultural heritage: an investigation into the feasibility of the OAIS model for application in small organisations. *Aslib Proceedings*. ISSN 0001-253X. Vol. 58, nº 6, (2006), p. 513-524.
- TAYLOR, Arlene G. (2004) – *The organization of information*. 2nd ed. London : Librarians Unlimited. ISBN 1-56308-976-9.
- THIBODEAU, Kenneth (2002) – Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In *The State of Digital Preservation : An International perspective : Conference Proceeding* [Em linha]. Washington, DC: Council on Library and Information Resources, 2002. [Acedido em 2010-03-06]. Disponível na WWW: <http://www.clir.org/pubs/reports/pub107/thibodeau.html>.
- TOMAZ, Katia P. ; SOARES, Antonio José (2004) – A preservação digital e o modelo de referência Open Archival Information System (OAIS) [Em linha]. *Datagramazero – Revista de Ciência da Informação*. Vol. 7, nº 1 (Fev. 2004). [Acedido em 2010-08-24]. Disponível na WWW: http://www.dgz.org.br/fev04/F_I_art.htm. ISSN 15173801.
- THOMAZ, Kátia P. (2005) – Gestão e preservação de documentos electrónicos de arquivo : revisão da literatura – parte I. *Arquivística*. ISSN 1808-4826. Vol. 1, nº 2 (Jul./Dez. 2005), p. 8-30
- UE (2001) – *Directiva 2001/29/CE do Parlamento Europeu e do Conselho de 22 de Maio de 2001 relativa à harmonização de certos aspectos do direito de autor e dos direitos conexos na sociedade da informação* [Em linha]. [S.l. : s.n.]. [Acedido em 2010-01-13]. Disponível na WWW: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:EN:NOT>
- UNESCO (2003) – *Charter on the preservation of digital heritage* [Em linha]. [S.l.] : Unesco. [Acedido em 2010-03-24]. Disponível na WWW: http://portal.unesco.org/ci/en/ev.php-URL_ID=13367&URL_DO=DO_TOPIC&URL_SECTION=201.html
- UNESCO (2002) – *Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archive* [Em linha]. [S.l.] : Unesco. [Acedido em 2010-07-26]. Disponível na WWW: <http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>
- von SUCHODOLETZ ; D., RECHERT, K., & van den DOBBELSTEEN, M. (2010) – Software archives as a vital base for digital preservation strategies [Em linha]. In *The 5th International Conference on Open Repositories Madrid Spain 6th - 9th July 2010*. [S.l. : s.n.]. [Acedido em 2010-03-20]. Disponível na WWW: <http://eprints.rclis.org/handle/10760/14732>
- WEBB, Colin (dir.) (2003) – *Guidelines for the preservation of digital heritage* [Em linha]. [S.l.] : Unesco, 2003. [Acedido em 2010-03-22]. Disponível na WWW: <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>
- WIPO (1979) – *Berne Convention for the Protection of Literary and Artistic Works* [Em linha]. [S.l.] : WIPO. [Acedido em 2010-01-13]: Disponível na WWW: http://www.wipo.int/export/sites/www/treaties/en/ip/berne/pdf/trtdocs_wo001.pdf

WIPO (1996) – *WIPO Copyright Treaty* [Em linha]. Geneva: WIPO. [Acedido em 2010-01-13]:
Disponível na WWW:
http://www.wipo.int/export/sites/www/treaties/en/ip/wct/pdf/trtdocs_wo033.pdf

YEO, Geoffrey (2010) – “Nothing is the same as something else” : significant properties and notions of identity and originality. *Archival Science*. ISSN 1573-7519. Vol. 10, nº 2, p. 85-116.

Índice de figuras

Figura 1: <i>Informação digital criada, capturada e replicada no mundo inteiro</i> . Fonte: Grantz (2008, p. 5).....	7
Figura 2: <i>Pirâmide da preservação</i> . Fonte: Caplan (2008, p. 9)	16
Figura 3: <i>Ciclos de vida da preservação tradicional e da preservação digital</i> . Fonte: BRTF (2010, p. 25 adaptado).....	19
Figura 4: <i>Densidade de informação vs Esperança de vida do suporte</i> . Fonte: Conway (1996, adaptado).....	27
Figura 5: <i>Cadeia de interpretação desde o nível físico ao nível conceptual</i> . Fonte: Ferreira (2009, p. 16.).....	33
Figura 6: <i>Recuperação de um objecto digital através de um emulador</i>	50
Figura 7: <i>Degradação do objecto digital ao longo de sucessivas migrações</i> . Fonte: Ferreira (2009, p. 30).....	56
Figura 8: <i>Migração a-pedido</i> . Fonte Ferreira (2009, p. 40)	56
Figura 9: <i>Migração distribuída baseada em Web Services</i> . Fonte: Ferreira (2009, p. 31).....	58
Figura 10 – <i>Entidades externas de um arquivo OAIS</i>	75
Figura 11- <i>Modelo funcional do OAIS</i> . Fonte: Ferreira (2006, p. 28).....	76
Figura 12: <i>Metadados de preservação</i>	81
Figura 13: <i>Diagrama esquemático para a selecção das estratégias de preservação digital</i> . Fonte: Lee et al. (2002, p. 103, adaptado).....	119
Figura 14: <i>Conteúdo de um Pacote de Informação, segundo o modelo de referência OAIS</i>	123
Figura 15: <i>Fluxo dos Pacotes de Informação</i>	124

Índice de tabelas

Tabela 1: <i>Acções e estratégias de preservação digital e suas indicações de uso</i>	69
Tabela 2: <i>Factores de sustentabilidade dos formatos</i> . Fonte: Library of Congress (2007)	113
Tabela 3: <i>Factores de qualidade e funcionalidade dos formatos</i> . Fonte: Library of Congress (2007)	114
Tabela 4: <i>Formatos adequados para preservação</i> . Fonte: Arquivo da Web Portuguesa (2010)	115
Tabela 5: <i>Análise da legislação dos países da UE: cópias para uso bibliotecário (natureza das obras)</i>	127
Tabela 6: <i>Análise da legislação dos países da UE: cópias para uso bibliotecário (tipo e número de cópias)</i>	128
Tabela 7: <i>Análise da legislação dos países da UE: cópias para uso bibliotecário (finalidade da cópia)</i>	129
Tabela 8: <i>Análise da legislação dos países da UE: ilusão das medidas tecnológicas de protecção</i>	130